



Universidad de San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Ingeniería Mecánica Industrial

**DISEÑO DE INVESTIGACIÓN PARA LA FORMULACIÓN DE MODELO DE DATOS PARA  
ESTIMAR LA HUELLA DE CARBONO EN LA CADENA DE PRODUCCIÓN DEL SECTOR  
CAFÉ EN GUATEMALA**

**Julio Cesar Rustrian Monterroso**

Asesorado por la Dra. Flor de Mayo González Miranda

Guatemala, abril de 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERÍA

**DISEÑO DE INVESTIGACIÓN PARA LA FORMULACIÓN DE MODELO DE DATOS PARA  
ESTIMAR LA HUELLA DE CARBONO EN LA CADENA DE PRODUCCIÓN DEL SECTOR  
CAFÉ EN GUATEMALA**

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA  
FACULTAD DE INGENIERÍA  
POR

**JULIO CESAR RUSTRIAN MONTERROSO**  
ASESORADO POR LA DRA. FLOR DE MAYO GONZÁLEZ MIRANDA

AL CONFERÍRSELE EL TÍTULO DE

**INGENIERO INDUSTRIAL**

GUATEMALA, ABRIL DE 2022

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA  
FACULTAD DE INGENIERÍA



**NÓMINA DE JUNTA DIRECTIVA**

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Kevin Vladimir Cruz Lorente
VOCAL V	Br. Fernando José Paz González
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

**TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO**

DECANO	Ing. Pedro Antonio Aguilar Polanco
EXAMINADOR	Ing. Edgar Darío Álvarez Cotí
EXAMINADORA	Inga. Mayra Saadeth Arreaza Martínez
EXAMINADORA	Inga. Aurelia Anabela Cordova Estrada
SECRETARIA	Ing. Lesbia Magalí Herrera López

## **HONORABLE TRIBUNAL EXAMINADOR**

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

**DISEÑO DE INVESTIGACIÓN PARA LA FORMULACIÓN DE MODELO DE DATOS PARA  
ESTIMAR LA HUELLA DE CARBONO EN LA CADENA DE PRODUCCIÓN DEL SECTOR  
CAFÉ EN GUATEMALA**

Tema que me fuera asignado por la Dirección de la Escuela de Estudios de Postgrado, con fecha enero 2022.

**Julio Cesar Rustrian Monterroso**



EEPFI-PP-0162-2022

Guatemala, 12 de enero de 2022

Director  
César Ernesto Urquizú Rodas  
Escuela Ingeniería Mecánica Industrial  
Presente.

Estimado Ing. Urquizú

Reciba un cordial saludo de la Escuela de Estudios de Postgrado de la Facultad de Ingeniería.

El propósito de la presente es para informarle que se ha revisado y aprobado el Diseño de Investigación titulado: **FORMULACIÓN DE MODELO DE DATOS PARA ANALIZAR LA HUELLA DE CARBONO EN LA CADENA DE PRODUCCIÓN DEL SECTOR CAFÉ EN GUATEMALA**, el cual se enmarca en la línea de investigación: **Análisis de datos - Análisis de datos**, presentado por el estudiante **Julio Cesar Rustrian Monterroso** carné número **201114376**, quien optó por la modalidad del "PROCESO DE GRADUACIÓN DE LOS ESTUDIANTES DE LA FACULTAD DE INGENIERÍA OPCIÓN ESTUDIOS DE POSTGRADO". Previo a culminar sus estudios en la Maestría en ARTES en Ingeniería Para La Industria Con Especialidad En Ciencias De La Computación.

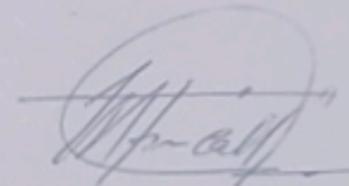
Y habiendo cumplido y aprobado con los requisitos establecidos en el normativo de este Proceso de Graduación en el Punto 6.2, aprobado por la Junta Directiva de la Facultad de Ingeniería en el Punto Décimo, Inciso 10.2 del Acta 28-2011 de fecha 19 de septiembre de 2011, firmo y sello la presente para el trámite correspondiente de graduación de Pregrado.

Atentamente,

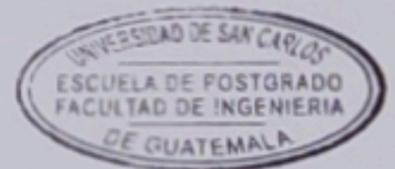
*"Id y Enseñad a Todos"*

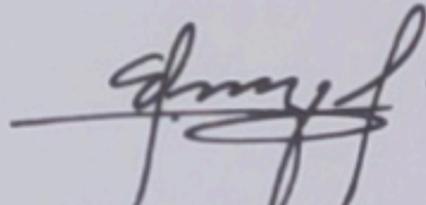
  
**Flor González Miranda**  
Ingeniera Industrial  
BSc. en Medio Ambiente  
Colegiada No. 6,070

Mtro. Flor De Mayo González Miranda  
Asesor(a)

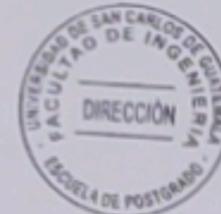


Mtro. Mario Renato Escobedo Martinez  
Coordinador(a) de Maestría





Mtro. Edgar Darío Álvarez Cotí  
Director  
Escuela de Estudios de Postgrado  
Facultad de Ingeniería





EEP-EIMI-0162-2022

El Director de la Escuela Ingeniería Mecánica Industrial de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del Asesor, el visto bueno del Coordinador y Director de la Escuela de Estudios de Postgrado, del Diseño de Investigación en la modalidad Estudios de Pregrado y Postgrado titulado: **FORMULACIÓN DE MODELO DE DATOS PARA ANALIZAR LA HUELLA DE CARBONO EN LA CADENA DE PRODUCCIÓN DEL SECTOR CAFÉ EN GUATEMALA**, presentado por el estudiante universitario **Julio Cesar Rustrian Monterroso**, procedo con el Aval del mismo, ya que cumple con los requisitos normados por la Facultad de Ingeniería en esta modalidad.

ID Y ENSEÑAD A TODOS

Ing. César Ernesto Urquizú Rodas  
Director  
Escuela Ingeniería Mecánica Industrial

Guatemala, enero de 2022

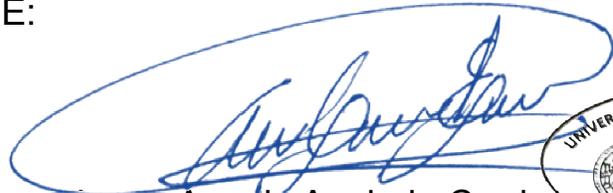


Decanato  
Facultad de Ingeniería  
24189101- 24189102  
secretariadecanato@ingenieria.usac.edu.gt

LNG.DECANATO.OI.230.2022

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería Mecánica Industrial, al Trabajo de Graduación titulado: **DISEÑO DE INVESTIGACIÓN PARA LA FORMULACIÓN DE MODELO DE DATOS PARA ESTIMAR LA HUELLA DE CARBONO EN LA CADENA DE PRODUCCIÓN DEL SECTOR CAFÉ EN GUATEMALA**, presentado por: **Julio Cesar Rustrian Monterroso**, después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, autoriza la impresión del mismo.

IMPRÍMASE:

  
Inga. Aurelia Anabela Cordova Estrada

Decana

Guatemala, abril de 2022

AACE/gaoc

## **ACTO QUE DEDICO A:**

### **Dios**

Por sobre todas las cosas, por guiarme, cuidarme y darme las fuerzas necesarias para poder seguir adelante durante todo el camino, por darme la oportunidad de finalizar con éxito esta etapa de mi vida y poner en mi camino a grandes personas que me apoyaron incondicionalmente.

### **Mis padres**

Juan Rustrian y María Monterroso de Rustrian por todo el apoyo, paciencia, cariño, por darme la oportunidad de desarrollarme como profesional y ser el motor que me impulsa día con día.

### **Mis hermanos**

Juan, José, y Sandra Rustrian por todo su apoyo, cariño, por estar allí cuando necesite de sus consejos y apoyo, y en especial agradezco a Walter y Hugo Rustrian por todo su sacrificio, apoyo, cariño y por ser un gran ejemplo en mi vida.

### **Mis sobrinos**

Jaqueline, Kevin, Diego, Katherine, Christopher y Evan Rustrian por su cariño, apoyo y por darme ánimos cuando lo he necesitado.

## **AGRADECIMIENTOS A:**

<b>Universidad de San Carlos de Guatemala</b>	Por abrirme las puertas hacía la educación superior y por brindarme la formación necesaria para ser un excelente profesional.
<b>Facultad de Ingeniería</b>	Por toda la enseñanza brindada, las noches de desvelo y por las amistades que me permitió hacer.
<b>Mis amigos de la Facultad</b>	María Serrano, Frank de la Roca, Luis Méndez, Jonathan Orantes, Andrea Argueta, Eduardo Paredes, entre otros, por ser una excelente compañía y grandes amigos durante toda esta experiencia.
<b>Mis catedráticos</b>	Por transmitir y compartir todo el conocimiento adquirido a lo largo de su experiencia con nosotros, para que el día de mañana seamos excelentes profesionales e Ingenieros.
<b>Mi asesora</b>	Dra. Flor de Mayo González por su apoyo y confianza durante esta etapa, por guiarme y darme la oportunidad de ampliar mis conocimientos en el área profesional.

## ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES .....	V
LISTA DE SÍMBOLOS .....	VII
GLOSARIO .....	IX
RESUMEN.....	XI
1. INTRODUCCIÓN .....	1
2. ANTECEDENTES .....	3
3. PLANTEAMIENTO DEL PROBLEMA .....	15
3.1. Contexto general.....	15
3.2. Descripción del problema .....	16
3.3. Formulación del problema .....	17
3.4. Delimitación del problema.....	18
4. JUSTIFICACIÓN .....	21
5. OBJETIVOS .....	23
5.1. General .....	23
5.2. Específicos.....	23
6. NECESIDADES A CUBRIR Y ESQUEMA DE LA SOLUCIÓN .....	25
7. MARCO TEÓRICO.....	27
7.1. SAF Café en Guatemala.....	27

7.2.	Unidad productiva.....	28
7.3.	Proceso productivo del SAF café .....	29
7.3.1.	Sistema agroforestal .....	30
7.3.2.	Beneficiado húmedo.....	31
7.3.3.	Beneficiado seco .....	33
7.3.4.	Tostaduría .....	34
7.3.5.	Gases de efecto invernadero (GEI).....	34
7.3.5.1.	Dióxido de carbono (CO <sub>2</sub> ) .....	36
7.3.5.2.	Captura y almacenamiento de CO <sub>2</sub> (CAC) .....	38
7.3.6.	Datos, información y conocimiento.....	40
7.3.6.1.	Datos.....	41
7.3.6.2.	Información.....	41
7.3.6.3.	Conocimiento .....	42
7.3.7.	Indicadores.....	43
7.3.8.	Estadística.....	45
7.3.8.1.	Medidas de tendencia central.....	45
7.3.8.2.	Medidas de dispersión .....	49
7.3.9.	Análisis de datos .....	52
7.3.10.	Bases de datos.....	52
7.3.11.	Técnicas de análisis de datos .....	54
7.3.12.	Arboles de decisión .....	54
7.3.12.1.	Aprendizaje automático.....	56
7.3.12.2.	Minería de datos.....	57
7.3.13.	Software de análisis estadístico .....	60
7.3.13.1.	Rstudio .....	61
8.	PROPUESTA DE ÍNDICE DE CONTENIDOS .....	63

9.	METODOLOGÍA.....	65
9.1.	Características del estudio.....	65
9.2.	Unidades de análisis.....	66
9.3.	VARIABLES.....	66
9.4.	Fases del estudio.....	68
9.4.1.	Revisión documental .....	68
9.4.2.	Recepción de información (datos crudos) .....	69
9.4.3.	Filtrado de información .....	69
9.4.4.	Análisis de información por medio de métodos estadísticos.....	70
9.4.5.	Clasificación de la información .....	70
9.4.6.	Desarrollo del modelo de datos .....	71
9.4.7.	Entrenamiento del modelo de datos .....	71
9.4.8.	Evaluación del modelo de datos.....	71
9.4.9.	Entrega final del modelo de datos .....	72
10.	TÉCNICAS DE ANÁLISIS DE INFORMACIÓN.....	73
10.1.	Preparación de la información .....	73
10.1.1.	Creación de la base de datos plana .....	73
10.1.2.	Carga de información al software de análisis .....	73
10.1.3.	Clasificación de la segmentación de la muestra .....	74
10.1.4.	Selección de tamaño de la muestra.....	74
10.1.5.	Error de muestreo.....	75
10.2.	Análisis de la información .....	75
10.2.1.	VARIABLES del proceso .....	76
10.2.2.	Análisis de tendencia central .....	77
10.2.3.	Análisis de dispersión .....	78
10.2.4.	Comparación entre variables.....	78
10.2.5.	Comportamiento entre variables.....	78

10.2.6.	Estudio del comportamiento de variables.....	79
10.3.	Presentación de los resultados.....	79
10.3.1.	Tablas de datos.....	80
10.3.2.	Gráficos de datos .....	81
11.	CRONOGRAMA .....	83
12.	FACTIBILIDAD DEL ESTUDIO .....	85
12.1.	Factibilidad operativa.....	85
12.2.	Factibilidad técnica .....	86
12.3.	Factibilidad económica .....	88
13.	REFERENCIAS .....	89

## ÍNDICE DE ILUSTRACIONES

### FIGURAS

1.	Unidades productivas de SAF café .....	29
2.	Sistema agroforestal de café.....	31
3.	Beneficiado húmedo de café .....	32
4.	Beneficiado seco de café .....	33
5.	Tostaduría .....	34
6.	Emisiones de GEI en EE.UU durante el año 2017 .....	36
7.	Emisiones de dióxido de carbono en EE.UU durante el año 2017 .....	38
8.	Sumideros de carbono .....	40
9.	Pirámide de conocimiento .....	43
10.	Representación gráfica de la media .....	47
11.	Árbol tipo ID3 .....	55
12.	Árbol tipo C4.5 .....	56
13.	Variables del proceso.....	77
14.	Ejemplo 1 de representación gráfica de datos .....	81
15.	Ejemplo 2 de representación gráfica de datos .....	82
16.	Cronograma de actividades propuestas.....	83

### TABLAS

I.	Clasificación de café Con base en altura de cosecha .....	27
II.	Regiones SAF Café.....	28
III.	Ejemplo de moda estadística .....	49
IV.	Variables de estudio.....	66

V.	Definiendo tamaño de la muestra .....	75
VI.	Ejemplo cantidad de CO2 emitido por unidad productiva .....	80
VII.	Ejemplo comparación de CO2 emitido por unidad productiva .....	81
VIII.	Datos de factibilidad económica .....	88

## LISTA DE SÍMBOLOS

<b>Símbolo</b>	<b>Significado</b>
<b>cm</b>	Centímetro
<b>CO<sub>2</sub></b>	Dióxido de carbono
<b>CO<sub>2e</sub></b>	Dióxido de carbono
<b>GB</b>	Gigabyte
<b>kg</b>	Kilogramo
<b>kW</b>	Kilovatio
<b>μ</b>	Media poblacional
<b>CH<sub>4</sub></b>	Metano
<b>m</b>	Metro
<b>m<sup>2</sup></b>	Metro cuadrado
<b>mm</b>	Milímetro
<b>nm</b>	Nanómetro
<b>N<sub>2</sub>O</b>	Óxido nitroso
<b>ft</b>	Pie
<b>qq</b>	Quintal
<b>Σ</b>	Suma
<b>TCO<sub>2</sub></b>	Toneladas de dióxido de carbono



## GLOSARIO

<b>Anacafé</b>	Siglas de Asociación nacional del café.
<b>Biomasa</b>	Materia de origen orgánico que se utiliza como fuente de energía renovable para distintos procesos.
<b>GEI</b>	Siglas de gases de efecto invernadero.
<b>Huella de carbón</b>	Es el conjunto de gases de efecto invernadero que deja toda actividad humana realizada.
<b>KPI</b>	Sigla de <i>key performance indicator</i> o también conocido como indicador tiene como función medir el rendimiento de un proceso.
<b>MIICC</b>	Sigla de maestría en ingeniería para la industria con especialización en ciencias de la computación.
<b>MP</b>	Siglas de materia prima.
<b>Phyton</b>	Lenguaje o <i>software</i> de programación utilizado para el desarrollo de otro software o para el análisis.
<b>PIB</b>	Sigla de producto interno bruto.
<b>Ratificar</b>	Palabra utilizada para aprobar o confirmar algo.

<b>SAF</b>	Sigla de sistema agroforestal.
<b>SCRUM</b>	Metodología utilizada para gestionar el desarrollo de software.
<b>Sumidero</b>	Proceso que se encarga de sustraer carbono de la atmosfera.
<b>SNM</b>	Sigla de sobre el nivel del mar.
<b>Software</b>	Programa de computación que permite llevar a cabo tareas en específico, el <i>software</i> varía en función de la tarea deseada.

## RESUMEN

En el presente trabajo de graduación se tiene como objetivo, proponer un modelo de datos el cual permita estimar la huella de carbono generada durante el proceso de la cadena de producción del sector café de Guatemala, tomando como base para el estudio tres diferentes áreas cafetaleras del sector.

Con el modelo propuesto se tiene un mejor control sobre los GEI generados, y se puede llevar a cabo la toma de decisiones o planteamiento de soluciones óptimas para mitigar la cantidad de CO<sub>2</sub> emitido.

Para llevar a cabo el modelo de datos, se realizó una serie de pasos ordenados estratégicamente, con el objetivo de obtener los mejores resultados posibles en el modelo. Para ello, fue fundamental establecer un periodo de tiempo para la selección, recolección y depuración de los datos para obtener una base de datos limpia y funcional; se procedió a entrenar el modelo con un conjunto de datos previo a desarrollar el modelo final y funcional con el cual se estimará la huella de carbono en la cadena de producción del sector café de Guatemala.



## 1. INTRODUCCIÓN

En la actualidad, a nivel mundial el sector empresarial se está dando a la tarea de ser más amigable con el medio ambiente en sus procesos de producción, con el objetivo de mitigar la cantidad de gases de efecto invernadero (GEI) y disminuir su huella de carbono generada, a lo cual, el sector cafetalero a nivel mundial no se ve exento y está optando por implementar acciones que le permitan llevar a cabo su producción emitiendo la menor cantidad de GEI's, es por esto que es de suma importancia lograr tener el control sobre la huella de carbono en el sector café de Guatemala.

Los datos como fuente de información son de vital importancia para mejorar la calidad y posicionamiento del café de Guatemala, actualmente se desconoce el impacto que generan las actividades como el riego del almacigo, renovación de plantaciones, nutrición del cultivo, despulpado, limpieza y trillado del café, por lo cual, es fundamental conocer el comportamiento y contribución que tiene cada una de estas etapas dentro la generación de CO<sub>2</sub>.

En el presente trabajo se propone llevar a cabo la formulación de un modelo de datos, que permita estimar la huella de carbono generada en la cadena de producción del sector café de Guatemala y a la vez permita observar el impacto que genera cada etapa dentro del mismo, esto beneficiara al sector café de Guatemala, permitiéndole: ser aún más productivo y agradable con el medio ambiente en sus diversas actividades de producción; La creación de nuevas KPI's de control por fase para la mejora continua; Creación de planes de mitigación y control; Brindar respuestas más eficientes a los caficultores para indicar donde se encuentra su mayor generación de CO<sub>2</sub>, entre otros. Gracias a

esta herramienta se podrá optar a una mayor presencia y reconocimiento internacional como a créditos de carbono.

Para iniciar este análisis de datos se estudiarán las variables de mayor impacto durante el proceso de producción del café por medio de herramientas de análisis de datos, software estadístico, entre otros, tomando como base de la investigación tres regiones cafetaleras establecidas con base en datos geográficos por la Asociación Nacional del Café – Anacafé -.

## 2. ANTECEDENTES

Anacafé (2016). En la política titulada *Política de Ambiente y Cambio Climático para el sector café de Guatemala*, se tiene como objetivo primordial el brindar los lineamientos adecuados estratégicamente para mejorar la gestión ambiental, la adaptación a los cambios en el clima, la competitividad y desarrollo con bajas emisiones de CO<sub>2</sub> dentro del marco de producción del sector café.

Se indican algunos principios básicos los cuales se mencionan a continuación:

- Sostenibilidad
- Corresponsabilidad equitativa
- Innovación y tecnología
- Mejora continua
- Adaptabilidad
- Constante formación de capacidades y participación

A la vez se indica los 4 ejes considerados de mayor importancia para esta política Adaptación y mitigación al cambio climático, Cumplimiento de la legislación ambiental nacional y de tratados internacionales ligados al sector café, Gestión integral de los recursos hídricos vinculados al sector café y Conservación y manejo sostenible de la biodiversidad en el sector café (Anacafé, 2016, p. 7).

Estos ejes son importantes, porque orientan indicando la forma más adecuada en que se debe realizar la producción en el SAF café de Guatemala, por otro lado, buscan que el sector café sea reconocido a nivel internacional, dando soporte utilizando guías para la mejora y desarrollo del sector. Gracias a

la información de este documento se podrá observar de una forma más clara y precisa las variables involucradas en cada fase de la producción y permitirán observar cuál de ellas causa un mayor impacto en la emisión de GEI.

Gavilán y Reinoso (2016). En la investigación denominada *Estimación cuantitativa de la huella del carbono en el cultivo de la caña de azúcar en villa clara*, desarrollado en la facultad de Ciencias Agropecuarias en la universidad Central Marta Abreu de las Villas, Cuba. Se puede observar que carecen del equipo adecuado para controlar el dióxido de carbono que se emite por la industria del país (haciendo énfasis en el sector cañero).

Dentro de este se indica que existen tres etapas esenciales que deben llevarse a cabo para su desarrollo, siendo estas: recolección de la información, análisis de la información y procesamiento de esta. A continuación, un ejemplo de herramientas que pueden ser utilizadas para el proceso:

- Listas de chequeo
- Entrevistas
- Observación directa
- Entre otros

Se hace mención del uso de un modelo matemático el cual consta de cuatro variables para su elaboración: (I) Impacto ambiental de la subhuella, (II) Toneladas generadas, (III) Factor de equivalencia y (IV) Factor de emisión.

Las variables de mayor impacto fueron identificadas en la producción de la caña de azúcar, y evidencio el cambio positivo en base al estudio realizado no solo a nivel medio ambiental sino a nivel administrativo, entre ellos, la reducción

de recursos utilizados, reduciendo implícitamente los costos de operación; y la mitigación de dióxido de carbono generada.

Con base en los resultados obtenidos se propuso llevar a cabo otras acciones similares, tomando como ejemplo los diversos procesos de la caña de azúcar: el reciclaje de los desechos, optimización de procesos como la cosecha y transporte, reducción del uso de combustible fósil entre otros.

Con base en lo expuesto, esta investigación aporta una metodología apropiada para dar inicio al análisis de la huella de carbón emitida en el proceso de producción del SAF café, ya que, aunque los estudios se aplican en diferentes áreas, existe una gran similitud entre ambos campos de estudio, y puede ser de gran valor técnico el uso de herramientas como listas de chequeos, entrevistas y un modelo matemático de justa aplicación.

Albornoz (2017). En la tesis titulada *Huella de Carbono del café (Coffea arabica) en Empresa Asociativa Campesina Aruco en Copán, Honduras para el año 2016-2017*, realizada en la escuela Agrícola Panamericana, Honduras.

Este trabajo nace de la necesidad de cumplir con los estándares medioambientales plasmados por el mercado extranjero, que cada vez se vuelven más exigentes, y al momento de intentar conocer el monto de las emisiones de gases, se observó que las unidades productivas no contaban con datos de sus respectivas actividades. Para este trabajo en específico se realizó la medición en tres diferentes estaciones del proceso de trabajo del café: cultivo, procesamiento y distribución.

Al momento de consolidar la base de datos con las respectivas cantidades de CO<sub>2</sub> emitidas durante cada etapa y con la ayuda de herramienta especializada

para el análisis de datos se consolida la cantidad de emisiones totales durante todo el sistema de producción.

Se concluyó lo siguiente con base en los resultados obtenidos:

- Para una cantidad de 1000 gramos de café tipo oro, el equivalente de dióxido de carbono corresponde a 3.42 kg en la producción y distribución en conjunto.
- Durante una medición de dióxido de carbono llevada a cabo durante el año 2011 en una finca de café, por cada mil gramos de café se obtuvo la cantidad de 2.78 kg de CO<sub>2e</sub>.
- Se observa que en la etapa conformada por la siembra y cosecha es donde se alcanza el pico máximo de emisión de gases, con una cantidad de 1.52 kg de CO<sub>2e</sub> debido al uso de nutrientes externos y energías de fuentes renovables.

Por lo cual a partir de los datos incluidos en el estudio y a las herramientas utilizadas para recopilación de los mismos, se puede inferir en que el estudio aporta una metodología apropiada realizar el estudio de dióxido de carbono emitido en el sector del café.

Mora y Mendoza (2017). En el trabajo nombrado *Determinación de impactos ambientales en la finca Cascajal en Pacho Cundinamarca, mediante el balance de carbono y nutrientes para un cultivo de café*, desarrollado en la Universidad de La Salle, Bogotá, busca determinar el impacto ambiental proveniente del proceso de la siembra y cosecha de café y determinar el impacto que causa el uso de fertilizantes en el mismo. Para llevar a cabo este estudio se planteó la división de la metodología en dos fases, las cuales se indican a continuación:

- Evaluación de matriz de impacto ambiental
- Determinación de cantidad de carbono utilizando software de análisis

A la vez, se observó que el secuestro y almacenamiento del carbono, se puede realizar por tres distintas técnicas:

- Procesos naturales como la fotosíntesis
- Técnicas de Ingeniería
- Transformaciones químicas

Como punto de apoyo, en este documento se hace mención a que se determinó que el uso de ciertos fertilizantes causa un impacto considerable dentro de la emisión de CO<sub>2</sub>, y se sugiere tener conocimiento sobre la biomasa del sistema para determinar la captación y almacenamiento de estas emisiones.

Este trabajo de graduación proporciona información importante al presente trabajo, ya que brinda observaciones respecto a variables importantes dentro del cálculo del dióxido de carbono emitido, da una guía sobre algunas fuentes importantes a tomar en consideración un acercamiento a una posible herramienta tecnológica para ratificar los datos.

Quirós (2017). En el artículo titulado *Inventarios de GEI y Huella de carbono para Beneficios*, se brindan información de carácter importante dando a conocer el origen de los gases denominados de efecto invernadero (GEI), así como algunas opciones de sumideros, propagación y control que permitan realizar un cambio en los GEI. Dentro del artículo, se puede observar el método utilizado para conocer rápidamente la cantidad de CO<sub>2</sub>, y la clasificación por cada etapa a la cual se somete el café.

Dentro de lo más sobresaliente del artículo, se puede mencionar la clasificación que brindan para el inventario de GEI y Huella de carbono, indicando los tres enfoques en que se debería clasificar cada contribución, siendo estos:

- Emisiones directas
- Emisiones indirectas y
- Emisiones indirectas externas

Contribuye al trabajo de investigación propuesto ya que brinda conocimiento relevante al mismo, como por ejemplo la clasificación de las emisiones, por otro lado, brinda soporte respecto a la idea de inventarios sencillos y funcionales.

Sánchez, Cabrera, Rojas, Ortiz, Gordillo y Perdomo (2017). En su artículo titulado *Software para el cálculo de la huella ambiental en la producción de cacao*, elaborado en la universidad cooperativa de Colombia, Colombia, se discute respecto a la importancia de la medición y el impacto que causa ambas huellas (hídrica y carbón) en la industria del cacao.

Este estudio sugiere el uso de distintas herramientas que permita la eficiencia del análisis de datos, utilizando el método SCRUM y analizando diversas etapas y variables o indicadores; así también se hace mención respecto a las normas de ambiente internacionales bajo las cuales se debe regir las actividades realizadas por un determinado individuo o grupo, y se indica se debe llevar a cabo bajo el marco establecido por las normas ISO 14064, PAS 2050, GHG Protocol, entre otras.

Con base en la información observada en este trabajo se logra distinguir las variables o indicadores importantes en la industria del cacao y su interacción

con la huella de carbono, por lo cual, se puede inferir en que será fuente importante de información para elaborar un análisis similar en el SAF café de Guatemala.

Ariza, Arias, Riaño, Riaño, Posada, Valenzuela, Vega, Murgueitio y Castro (2018). En el artículo titulado *Determinación de la huella de carbono en el sistema de producción de café pergamino seco de cuatro municipios del sur del departamento del Huila (Colombia)*, cuyo estudio se encarga de determinar la cantidad de CO<sub>2e</sub> emitido en la cadena de producción de café; el porcentaje correspondiente a la emisión de GEI. Con base en estos datos, procedieron a la generación y medición de indicadores que permitiera mejorar el análisis de datos y la toma de decisiones.

Con base en lo descrito en la Norma Técnica Colombiana NTC-5947 se desarrolló la metodología utilizada, tomando en consideración los distintos valores establecidos de huella de carbón, producción de GEI y su respectiva unidad funcional. Para dicho estudio se utilizaron modelos matemáticos no lineales basados en tiempos cronológicos y convirtiendo las unidades de medidas a la unidad de medida central siendo este dióxido de carbono equivalente (CO<sub>2eq</sub>).

Como factor clave y que contribuye de gran manera al trabajo de graduación actual en la búsqueda de calcular la huella de carbón que proviene de la producción del café de Guatemala es importante resaltar que, gracias a los resultados del estudio, se conoce que el diámetro del tallo de la planta de café a una distancia de 15 centímetros respecto al suelo es la mejor opción para la toma del muestreo de biomasa en el café.

Zaldaña (2019). En el trabajo *Determinación de emisiones de gases de efecto invernadero y alternativas para la gestión eficiente de captura de carbono*

*a través de la herramienta Cool Farm Tool en los sistemas de producción de Coffea sp. En Rodríguez de Mendoza – Amazonas 2018, realizado en la Universidad Nacional Toribio Rodríguez de Mendoza, Perú.*

En la búsqueda de optimizar los procesos de producción, mitigar los GEI emitidos y ser más agradable con el medio ambiente en el Perú, por medio del análisis obtenido con el uso de la herramienta Cool Farm, se realizó el estudio para determinar cuánto GEI se genera en este sector del café.

La metodología empleada para realizar dicha investigación se basó en un modelo experimental de tipo analítico denominado 2A\*3B, donde A representa el manejo de sombra (con o sin sombra), mientras que B representa el tipo de manejo que se llevaba a cabo (fertilización orgánica, fertilización química o sin fertilización).

El trabajo de investigación, así como la metodología y procesos realizados en el trabajo de graduación anteriormente mencionado puede ser una fuente de información de gran utilidad al actual trabajo de graduación, ya que se involucra temas correspondientes a diferentes sistemas de cultivo, factores de emisión de CO<sub>2</sub>, captación y secuestro de CO<sub>2</sub>, entre otros.

Olmos (2021). En el trabajo presentado como *Guía básica para la recolección y validación de los datos necesarios para calcular la huella ambiental del café verde según la norma europea*, se desarrolló el análisis de las etapas o fases para el sistema de producción de café y a la vez conocer la cantidad de huella de carbón que esta emite para el sector de Latinoamérica y el caribe.

Este documento tiene como fin el mostrar cómo llevar a cabo la correcta recopilación y validación de datos básicos para el cómputo de la huella de carbón;

se indican posibles mecanismos a utilizar, y se sugiere el uso de bibliografías que complementen esta información. La metodología se encuentra fundamentada en la guía de *Product Environmental Footprint*.

Según este documento, se debe tener conocimiento de que la huella ambiental es un indicador para un producto y que este mide el impacto que causa durante todo su ciclo vital, a la vez, hay que considerar todas las fases a la que el producto se somete:

- Producción de la MP
- Transporte
- Transformación
- Procesamiento
- Empaque
- Limpieza
- Uso del producto
- Disposición final

Dentro de este documento, es importante reconocer los datos brindados para determinar la correcta clasificación y medición de CO<sub>2</sub> en las fases que corresponden al ciclo del SAF café, tomando como ejemplo:

- Proceso de despulpe
- Proceso de lavado
- Proceso de secado
- Proceso de trillado
- Entre otros.

La metodología propuesta dentro de este documento *Product Environmental Footprint* será analizada para poder aplicarla al trabajo del cálculo del dióxido de carbono causado debido a la producción del café de Guatemala, también se verificará la información proveniente de los distintos procesos que se mencionan en el mismo para determinar cuáles serán fundamentales en este estudio.

Aguirre y Garay (2021). En la investigación titulada *Determinación de la huella de carbono en el cultivo y procesamiento del café y estrategias para su reducción*, desarrollado en la facultad de Ciencias Ambientales en la universidad Internacional SEK, Quito, Ecuador; Este trabajo se realizó para sugerir planes o estrategias de trabajo que permitieran reducir el dióxido de carbono generado debido a la producción del café.

Estas acciones se desarrollaron específicamente en una empresa dedicada al proceso productivo de café, la cual no cuenta con documentación o informes que muestren información correspondiente a las GEI durante el proceso productivo de este.

Por esto, por medio de metodología ya estipulada por Norma ISO 14064-1:2018 se procedió a realizar un inventario de GEI's donde se dieron a conocer los procesos establecidos para la producción del café.

Gracias a este estudio realizado se logró determinar que la empresa emitió durante el año 2019 un total de 21,663 tCO<sub>2</sub>. Esta cantidad de GEI's se distribuye dentro de distintos procesos los cuales se describirán a continuación:

- Transporte por vía aérea (Exportación) equivalente a un 77.46 %
- Cambio de uso de suelo equivalente a un 11.75 %

- Las emisiones del uso de combustibles corresponden a un 3.91 %
- Uso de fertilizantes nitrogenados un 3 %
- Disposición y uso de residuos sólidos para compostaje un 2.59 %
- Utilización del compostaje como fertilizante un 0.62 %
- Otras actividades realizadas un 0.67 %

Con base en esta información se propuso un plan, el cual contiene diversas medidas de mitigación que ayudarán tanto en la rentabilidad del negocio, como podrán optar a la compra de bonos de carbón, y contribuirán con la mitigación de los GEI enviados a la atmósfera.

Todo el contenido correspondiente a este proyecto aporta una metodología apropiada desarrollar el estudio de la cantidad de carbón generado en la producción del café de Guatemala, ya que, existe similitudes tanto en el tema central, como en la carencia de herramientas, y podría llegar a ser una gran base informativa para llevar a cabo el desarrollo de este trabajo de graduación.



### **3. PLANTEAMIENTO DEL PROBLEMA**

#### **3.1. Contexto general**

Guatemala cuenta con una gran diversidad de climas y zonas que le permiten cultivar café de gran calidad, es por esto que el sector café de Guatemala pese a contar en su mayoría con micro y pequeños caficultores, es una de las industrias que permite generar una gran cantidad de empleos en sus diversas actividades, así como brinda un gran aporte al producto interno bruto (PIB) del país.

A nivel nacional se cuenta con diversas unidades productivas y sistemas agroforestales, en los cuales se lleva a cabo el cultivo y producción de SAF café para consumo tanto a nivel nacional como para la exportación hacia distintos países del mundo.

Actualmente la Asociación Nacional del Café – Anacafé- es la institución que apoya al sector cafetalero de Guatemala y se encarga de velar por la mejora continua en todos los procesos correspondientes a la siembra, cosecha y producción de café. El proyecto de producción con bajas emisiones de gases de efecto invernadero, se ha llevado a cabo en algunos sectores productivos del territorio guatemalteco, sin embargo, aún no se ha implementado por completo dentro del sector SAF café. La falta de información que motive al sector productivo de Guatemala a lograr una producción con bajas emisiones GEI, es clave para dar inicio al desarrollo sostenible de nuestras empresas. Hablar de productividad con bajas emisiones, mitigación de la huella de carbono u otros conceptos similares referentes a llevar a cabo nuestra producción de una forma

más amigable con el medio ambiente, es aún un tema que se encuentra en pleno crecimiento a nivel mundial.

Gracias a un estudio llevado a cabo en Guatemala durante el año 2016, se logró identificar ciertos aspectos que actúan como barreras al momento de llevar a cabo el cálculo de emisiones de gases de efecto invernadero, de entre las cuales se puede mencionar:

- La falta de herramientas y sistemas de monitoreo,
- Falta de capacidad o consultoría especializada para el control de datos,
- Desconocimiento de ventajas competitivas y operativas,
- Entre otros.

### **3.2. Descripción del problema**

Anacafé actualmente no cuenta con una herramienta o modelo implementado que le permita llevar el control de la cantidad de dióxido de carbono (CO<sub>2</sub>) que se genera en las distintas unidades productivas de SAF café, por lo cual, sin esta herramienta no puede brindar el soporte adecuado según sea la necesidad por cubrir. Se debe llevar a cabo un análisis estadístico/matemático en el cual se involucren variables de alto impacto en la emisión de CO<sub>2</sub> y variables neutras que, aunque no serán generadoras de CO<sub>2</sub> forman parte importante en el proceso de captación de este.

Una vez establecidas las variables anteriormente indicadas, se puede llevar a cabo un modelo en el cual se pueda estimar la huella de carbono generada. Actualmente se puede identificar algunos factores de impacto ambiental dentro de la huella de carbono generada, siendo estos:

- Combustibles
- Energía eléctrica
- Insumos agrícolas

Al no contar con un control adecuado de la huella de carbono causada por estos factores, se puede incurrir en algunos efectos o consecuencias futuras como podría ser: menor reconocimiento a nivel internacional; daño al suelo de cultivo, disminuyendo la capacidad de absorción de nutrientes y minerales, y con esto la disminución en la capacidad de producción; El aumento en las emisiones de gases de efecto invernadero (GEI).

### **3.3. Formulación del problema**

A continuación, se presentan las preguntas necesarias para el desarrollo de esta investigación.

- Pregunta central

¿Cómo determinar la cantidad de huella de carbono que se genera en la cadena de producción del SAF café de Guatemala?

- Preguntas auxiliares
  - ¿Qué porcentaje de pulpa se está generando durante el proceso productivo del SAF café?
  - ¿Cuántos galones de combustible se utilizan durante las distintas actividades de producción?
  - ¿Cuántos metros cuadrados [m<sup>2</sup>] de sombra se encuentran en el sistema agroforestal?

- ¿Qué cantidad de energía eléctrica mensual [KW/mes] es utilizada durante el proceso productivo del SAF café?
- ¿Qué cantidad de insumos agrícolas se utiliza al mes durante el proceso del sistema agroforestal del café?

### **3.4. Delimitación del problema**

El sector café de Guatemala es muy amplio y la actividad productiva se desarrolla a lo largo de los 22 departamentos y en 204 de los 340 municipios del país. Anacafé tiene cubiertas 8 regiones cafetaleras a nivel nacional, siendo estas:

- Acatenango
- Antigua Guatemala
- Fraijanes
- Atitlán
- Huehuetenango
- Cobán
- San Marcos
- Nuevo Oriente

El análisis de datos correspondiente al presente trabajo se elaborará específicamente para las áreas de:

- Acatenango
- Antigua Guatemala
- Fraijanes

Tomando como referencia los datos correspondientes al periodo 2017 y 2022.



## 4. JUSTIFICACIÓN

La elaboración del presente trabajo se relaciona con la línea de investigación de Análisis de Datos para el grado de postgrado, correspondiente a la maestría para la industria con especialización en ciencias de la computación –MIICC-.

El SAF café de Guatemala es una de las piezas claves de nuestra economía local, provee fuentes de trabajo en los distintos departamentos del país, así como también es reconocido a nivel internacional, siendo este uno de los diez (10) mayores productores de café en el mundo. Actualmente, se está ante una etapa social a la cual se puede denominar revolución ecológica, la cual está obligando a distintos sectores de producción a ser más amigables con el medio ambiente. Tener un control adecuado de la huella de carbono producida por la cadena de producción del SAF café, permitirá reducir los gases de efecto invernadero (GEI) producidos y como efecto cadena mitigar el impacto ambiental, implícitamente se podrá obtener una reducción en los costos de la empresa y puede ser la diferencia clave entre destacar o atrasarse ante los mejores productores de café del mundo.

Se plantea llevar a cabo un modelo matemático, el cual permita estimar la huella de carbono y poder tener un control adecuado de la misma, de esta forma tanto los productores como el sector SAF café en general se verá beneficiado y podrá apoyarse con una herramienta con la cual puedan llevar a cabo acciones o plantear soluciones que permitan mitigar la cantidad de dióxido de carbono emitido.

Esta investigación contribuirá con una mejorara sustancial al sector SAF café de Guatemala, permitiendo la producción de un café de mayor calidad y prestigio, se proveerá de un nuevo indicador fácil de entender para cálculos futuros relacionados con temas ambientales y se contara con una ventaja competitiva a nivel internacional.

## 5. OBJETIVOS

### 5.1. General

Diseñar un modelo para la estimación de la huella de carbono en la cadena de producción del SAF café de Guatemala.

### 5.2. Específicos

- Estimar la cantidad de CO<sub>2</sub> emitida por la gestión de la pulpa durante el proceso productivo del SAF café.
- Determinar cuántos galones de combustible se utilizan durante las distintas actividades de la cadena de producción de SAF café y su estimado en CO<sub>2</sub>.
- Identificar la cantidad de metros cuadrados [m<sup>2</sup>] de sombra que se encuentran en el sistema agroforestal y su equivalente a captación de CO<sub>2</sub>.
- Mostrar el importe de energía eléctrica mensual [KW/año] que se utiliza durante el proceso productivo del SAF café y su equivalente en emisión de CO<sub>2</sub>.
- Registrar la cantidad de insumos agrícolas que se utilizan al mes durante el proceso del sistema agroforestal del café y su generación estimada de CO<sub>2</sub>.



## **6. NECESIDADES POR CUBRIR Y ESQUEMA DE LA SOLUCIÓN**

Guatemala es uno de los principales productores y exportadores de café a nivel mundial, produciendo café de renombre y gran calidad, sin embargo, en la actualidad aún no se cuenta con una herramienta que permita tener el control de la huella de carbono emitida durante el proceso de producción del mismo, eso representa un descontrol en la cantidad de emisiones de Gases de Efecto Invernadero (GEI) lo cual pone en desventaja al producto nacional respecto a productos internacionales que día con día se encuentran en la búsqueda de producto más amigable con el medio ambiente. La ausencia del control de emisiones GEI también repercute en mayores costos de producción y costo de oportunidad ante posibles convenios con otros países.

El modelo propuesto en este documento se enfoca en desarrollar una herramienta que permita al sector cafetalero de Guatemala poder estimar la cantidad de huella de carbono que se genera durante el proceso de producción del SAF café, tomando en cuenta las diversas actividades que se desarrollan durante el mismo, tanto desde la siembra y cosecha del SAF café hasta la etapa de trillado. Esta herramienta buscara incentivar a las diversas unidades productivas del país a iniciar con un sistema de producción más amigable con el medio ambiente y optar por incentivos internacionales como por ejemplo bonos de carbono.

Para poder realizar este modelo se deberá llevar a cabo una serie de actividades como se describe a continuación.

Una vez elegidas las fuentes de información y las áreas donde se llevará a cabo el trabajo de investigación, se elegirá el periodo de tiempo durante el cual se llevará a cabo el análisis de la huella de carbono; Con las variables de trabajo identificadas, se procederá a la recolección de datos por cada una de las variables de estudio y se generará una base de datos con la cual se iniciará la elaboración del modelo.

Se llevará a cabo la depuración de datos para consolidar la base de estos.

Una vez la base de datos se encuentre consolidada, se dará inicio al estudio de los datos, verificando la correlación entre las distintas variables involucradas.

Se procederá a identificar la cantidad de huella de carbono equivalente en función de la unidad de medida de cada variable analizada, investigando y basándose en fuentes oficiales reconocidas y se tomará nota de esta para futura referencia. Una vez se ha llevado a cabo la equivalencia de unidades de medida, se deberá clasificar los datos en función de su relevancia dentro del proceso, esto por medio de una calificación o ponderación específica.

Una vez se tiene preparado todo lo anteriormente mencionado, se continuará con la siguiente actividad la cual involucra realizar los respectivos cálculos y brindar el producto final del modelo.

## 7. MARCO TEÓRICO

### 7.1. SAF Café en Guatemala

Guatemala, es uno de los países natos en la siembra y cosecha de café a nivel mundial, esto gracias a la gran diversidad de climas y área geográfica con los que se cuenta, pero desde 1859 es cuando toma mayor auge dentro del sector productivo del país y es en este mismo año cuando se da inicio a la exportación del café. El inicio de estas exportaciones colocó al café como uno de los productos con mayor relevancia dentro la producción del país ya que este abarcaba casi el 80 % de las exportaciones.

Guatemala cuenta con más de 305 hectáreas de SAF café distribuidas a lo largo del país, esto representa un 2.8 % del territorio nacional, cuenta con cultivos alrededor de 22 departamentos y en 204 de los 340 municipios.

Actualmente el sector cafetalero de Guatemala se basa en la producción de SAF café arábigo, el cual Con base en una determinada clasificación según la altura del sembrado se subdivide en:

Tabla I. **Clasificación de café Con base en altura de cosecha**

<b>Rango de altura [ft snm]</b>	<b>Subdivisión</b>
2,500 – 3,000	Prima lavado
3,000 – 3,500	Extra prima lavado
3,500 – 4,000	Semiduro
4,000 – 4,500	Duro
> 4,500	Estrictamente duro

Fuente: elaboración propia, empleando Microsoft Excel 365.

En el año 2019 el sistema agroforestal de café abarco aproximadamente el 7.32 % de las exportaciones a nivel nacional, lo cual lo clasifica como el segundo producto de mayor exportación por debajo del área textil. En la cosecha correspondiente al periodo octubre 2019-septiembre 2020 la exportación de café se sostuvo pese a la problemática marcada por el COVID-19, alcanzando los 4.16 millones de quintales de café en su variante de café oro, con lo cual estas exportaciones contribuyeron con un total de Q 5,111.6 millones en divisas.

Actualmente, Anacafé clasifica las unidades productivas en siete distintas regiones, basándose en su ubicación geográfica, a continuación, se describe esta clasificación:

Tabla II. **Regiones SAF café**

#Región	Departamentos
I	San marcos y Quetzaltenango
II	Sololá, Suchitepéquez, Retalhuleu, El Palmar (Quetzaltenango) y Pochuta (Chimaltenango)
III	Guatemala, Chimaltenango, El Progreso, Escuintla y Sacatepéquez
IV	Jutiapa, Santa Rosa y Jalapa
V	Guatemala, Chimaltenango, El Progreso, Escuintla y Sacatepéquez
VI	Alta y Baja Verapaz y El Estor Izabal
VII	Zacapa, Chiquimula y Morales Izabal

Fuente: elaboración propia, empleando Microsoft Excel 365.

## 7.2. Unidad productiva

Se denomina como unidad productiva de SAF café, a todos aquellos actores involucrados en el desarrollo de la actividad de producción de SAF café, es decir, desde pequeños productores o productores individuales que no cuentan con grandes extensiones de terreno para su cosecha, hasta organizaciones de

productores de café asociados que cuentan con grandes extensiones de terreno para la siembra y cosecha de SAF café, como, por ejemplo, cooperativas o asociaciones.

Como es posible apreciar en la siguiente tabla, existen diversas unidades productivas distribuidas alrededor de las siete regiones cafetaleras definidas por Anacafé; al año 2020, 1,469 unidades productivas son individuales, mientras que 28,698 son unidades productivas de caficultores organizados, totalizando 30,167 unidades productivas en el territorio guatemalteco.

Figura 1. **Unidades productivas de SAF café**

Región	Productores			
	Individuales	Organizados	Total	Organizaciones
I	157	3,837	3,944	111
II	150	3,726	3,876	124
III	242	2,949	3,191	85
IV	308	1,983	2,291	47
V	137	7,171	7,308	172
VI	217	5,556	5,773	135
VII	258	3,476	3,734	74
<b>TOTAL</b>	<b>1,469</b>	<b>28,698</b>	<b>30,167</b>	<b>748</b>

Fuente: Anacafé. (2020). *Memoria de labores 2019 – 2020*.

### 7.3. Proceso productivo del SAF café

Entre la producción del SAF café se encuentran distintas etapas o actividades involucradas en la transformación de este, dentro de estas se puede mencionar el sistema agroforestal en fase de cosecha, fase de beneficiado húmedo y beneficiado seco, trillado y tostaduría. Cabe resaltar que no todas las unidades productivas tienen la factibilidad económica de contar con todas las

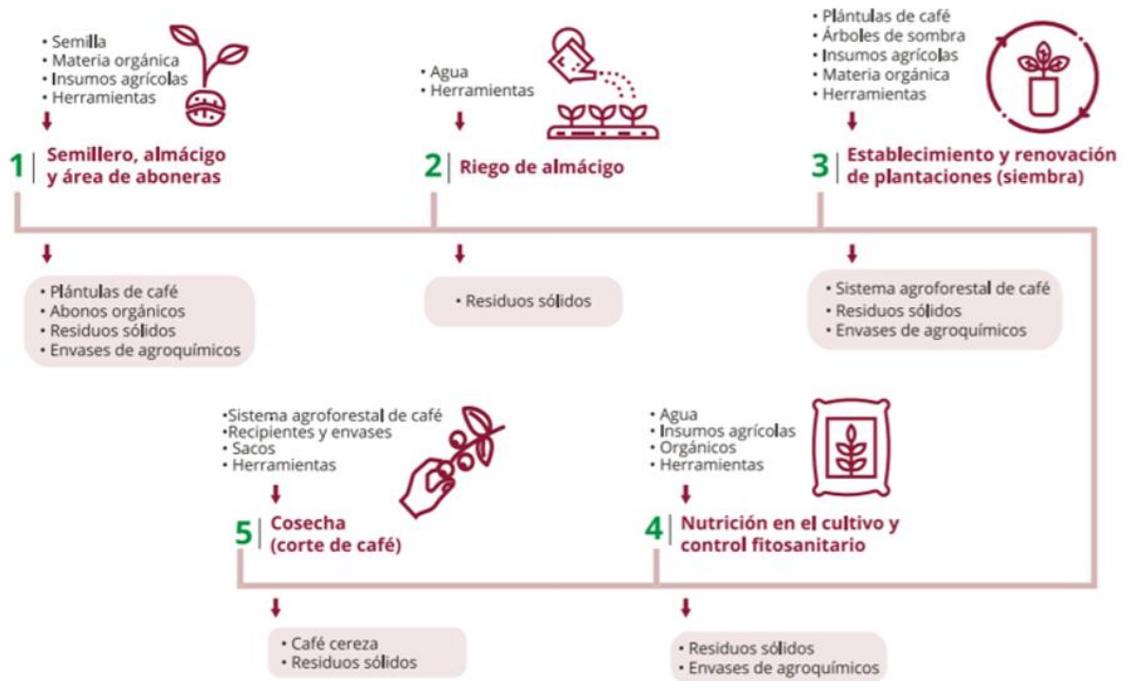
fases previamente indicadas y deben optar por acudir a servicios externos que brinden este servicio. Cada fase por individual puede emitir dióxido de carbono debido a algún proceso interno, y esta cantidad por poca que sea, debe contabilizarse para poder tener una correcta medición y control de la emisión de GEI.

### **7.3.1. Sistema agroforestal**

Puede denominarse como, conjunto de recursos naturales (árboles y plantas forestales) que son utilizados junto con otros cultivos de origen agrícola y en determinados casos también con distintas especies de animales. Esta unión de los recursos previamente mencionados permite diversificar la producción, aumentar el nivel de materia orgánica en el suelo de cultivo, reciclar nutrientes del suelo, captación de dióxido de carbono (CO<sub>2</sub>), entre otros.

Figura 2. Sistema agroforestal de café

## Sistema agroforestal de café



Fuente: Anacafé. (2019). *Guía ambiental para el sector café de Guatemala*.

### 7.3.2. Beneficiado húmedo

El beneficiado húmedo del SAF café, es el proceso durante el cual el café sufre la transformación desde la etapa inicial de café cereza cortado a la etapa final en la cual el resultado es el café pergamino seco, todo este proceso se ve apoyado por el uso de agua dentro de las etapas de este, dentro de las etapas de este proceso de transformación se encuentra:

- Recolección del grano.
- Despulpado, por medio de maquinaria y equipo con la ayuda de agua limpia o recirculada.
- Remoción de mucílago y lavado, en el cual nuevamente es necesario el uso de maquinaria y equipo en conjunto con combustible y/o electricidad y agua limpia o recirculada.
- Secamiento, en la cual se utiliza infraestructura e insumos de secado, herramientas y elementos de almacenamiento donde se coloca el café pergamino seco.

Figura 3. **Beneficiado húmedo de café**

## Beneficiado húmedo



Fuente: Anacafé (2019). *Guía ambiental para el sector café de Guatemala.*

### 7.3.3. Beneficiado seco

Son todas aquellas operaciones que se realizan para la transformación del café cereza en café pergamino seco sin necesidad de utilizar agua en sus procesos o estaciones de trabajo, dentro de las etapas de transformación se encuentra:

- Recolección del grano.
- Limpieza del café, por medio de maquinaria y equipo en conjunto con combustible y/o electricidad.
- Trillado y pulido, con el uso de maquinaria.
- Clasificación del grano.

Figura 4. Beneficiado seco de café

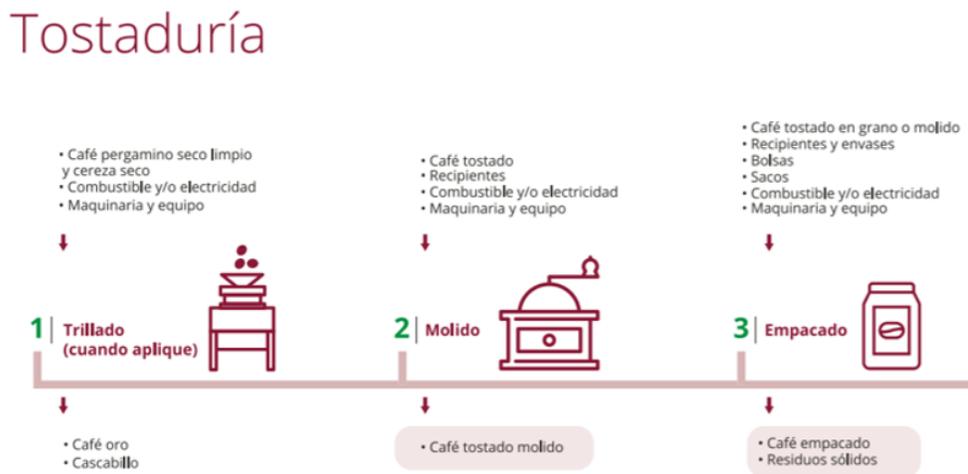


Fuente: Anacafé. (2019). *Guía ambiental para el sector café de Guatemala*.

### 7.3.4. Tostaduría

La etapa de tostaduría, es una fase de transformación que involucra café de tipo pergamino y con la ayuda de maquinaria, combustibles y agua, se remueve la cascara al grano dando como resultado final el grano de café tipo oro.

Figura 5. Tostaduría



Fuente: Anacafé. (2019). *Guía ambiental para el sector café de Guatemala*.

### 7.3.5. Gases de efecto invernadero (GEI)

Se conoce como gases de efecto invernadero (también conocidos por sus siglas GEI) a todos aquellos gases que son emitidos hacia la atmósfera, causando el efecto conocido como efecto invernadero. Estos GEI provienen de distintos orígenes como actividades desarrolladas por el ser humano, como de las herramientas o equipos que este utiliza para realizar dichas actividades, como ejemplo de los GEI se puede mencionar:

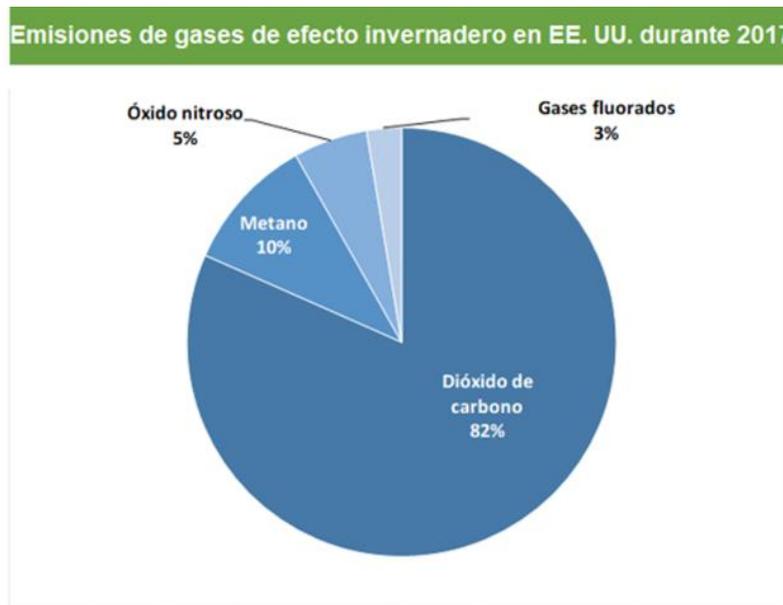
- Dióxido de carbono (CO<sub>2</sub>)
- Óxido nitroso (N<sub>2</sub>O)
- Gases fluorados
- Metano (CH<sub>4</sub>).

Y dentro de las actividades que pueden generar estos GEI se puede hacer mención a:

- Quema de combustibles fósiles
- Reacciones químicas
- Descomposición de residuos
- Tratamiento de aguas residuales
- Producción de gas natural
- Entre otros.

Cada uno de estos gases emitidos tiene diferente magnitud en la forma en que afectan a la atmosfera terrestre, en el siguiente gráfico es posible observar el comportamiento que tiene cada gas respecto a la totalidad de GEI emitidos en Estados Unidos en el año 2017.

Figura 6. **Emisiones de GEI en EE. UU durante el año 2017**



Fuente: Agencia de protección ambiental de Estados Unidos. (2021). *Emisiones de gases de efecto invernadero en EE-UU durante 2017*. Consultado el 25 de septiembre de 2021. Recuperado de <https://espanol.epa.gov/la-energia-y-el-medioambiente/descripcion-general-de-los-gases-de-efecto-invernadero>.

El gas que causa mayor impacto como GEI es el dióxido de carbono con un 82 %, seguido del Metano con un 10 %, el Óxido de nitrógeno con 5 % y por último, pero no menos despreciable los gases Fluorados con un 3 %.

### 7.3.5.1. Dióxido de carbono (CO<sub>2</sub>)

Gas natural que se encuentra presente en superficie de la atmosfera, y se emite por medio de diversas actividades o procesos, aumentando así la cantidad de GEI que se elevan hacia la atmosfera terrestre, este puede incluso llegar a producirse por el cuerpo humano al realizar sus respectivas actividades cotidianas.

Dentro de las actividades o procesos que generan CO<sub>2</sub> se pueden indicar:

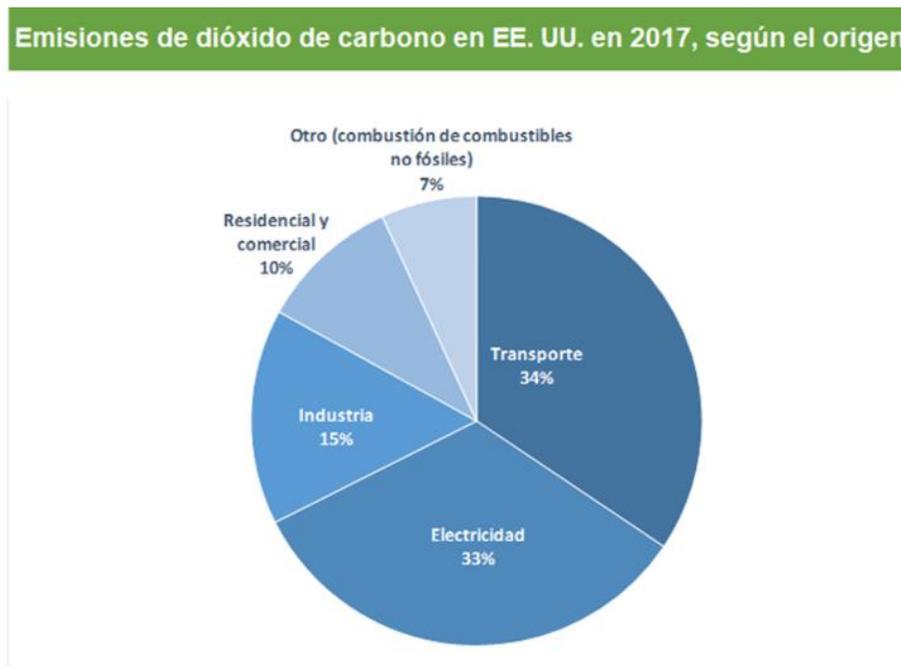
- Uso de combustibles fósiles
  - Carbón
  - Gas natural
  - petróleo
- Residuos sólidos
- Materiales biológicos
- Reacciones químicas

Según datos de la Agencia de Protección Ambiental de Estado Unidos (APA) durante el año 2017 en Estados Unidos las fuentes de emisión más grandes de CO<sub>2</sub> fueron las siguientes:

- Transporte
- Electricidad
- Industria (Procesos industriales)

El dióxido de carbono se resalta como el GEI más común y el de mayor influencia en el impacto ambiental del efecto invernadero generado.

Figura 7. **Emisiones de dióxido de carbono en EE. UU durante el año 2017**



Fuente: Agencia de protección ambiental de Estados Unidos (2021). *Emisiones de dióxido de carbono en EE-UU en 2017*. Consultado el 25 de septiembre de 2021. Recuperado de <https://espanol.epa.gov/la-energia-y-el-medioambiente/emisiones-de-dioxido-de-carbono>.

### 7.3.5.2. **Captura y almacenamiento de CO<sub>2</sub> (CAC)**

El proceso de capturar o secuestrar el CO<sub>2</sub> y posteriormente almacenarlo, es una de las técnicas que actualmente se puede considerar como una de las opciones viables en la búsqueda de acciones que permitan mitigar o estabilizar la cantidad de GEI que se emiten dentro de diversos sectores de la industria y sus procesos de producción. Gracias a técnicas como la captura y almacenamiento de CO<sub>2</sub> uno de los mayores beneficios podría ser de tipo medioambiental, permitiendo formar parte en la lucha de la mitigación de GEI

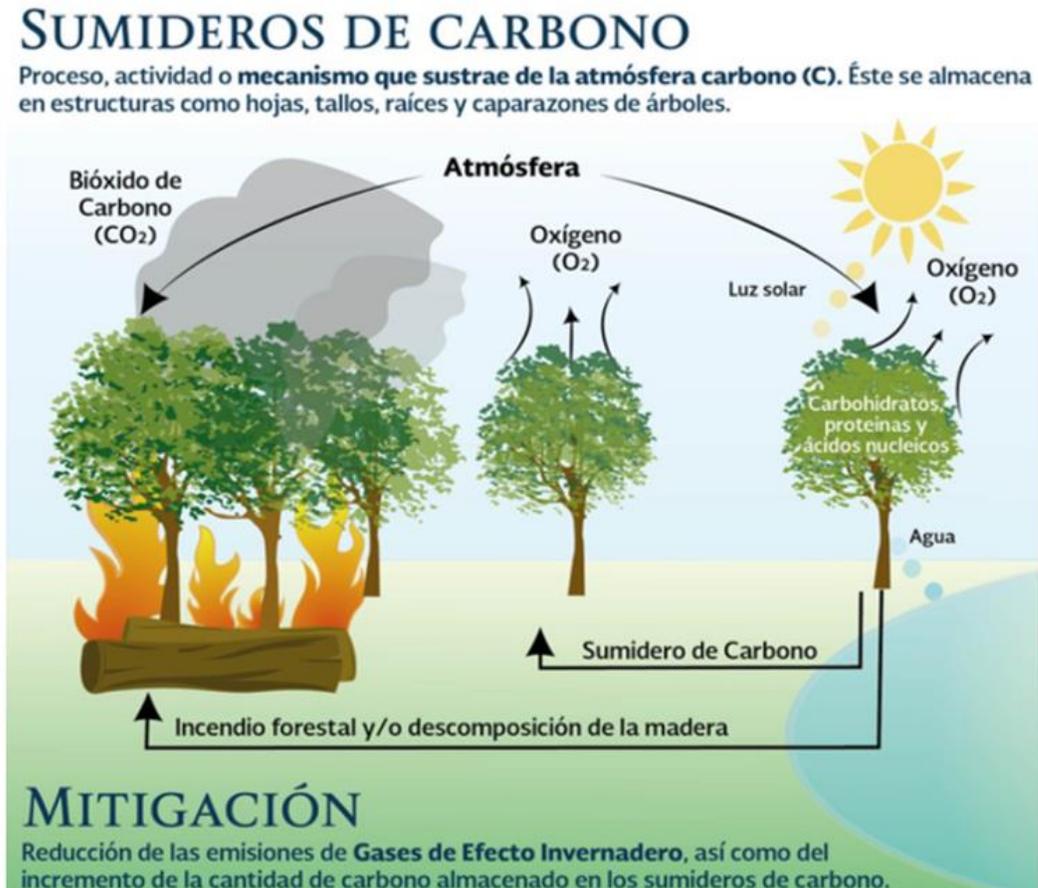
emitidos y estabilizando todos estos gases que se acumulan en la atmósfera terrestre, las empresas también podrán observar u obtener otros beneficios como la reducción en los costos operativos (mejorar el rendimiento energético, cambios en el tipo de combustible, optar por uso de energías renovables, entre otros) y a la vez, el reconocimiento y valoración tanto nacional como internacional dentro de un mercado verde y comprometido con el medio ambiente.

El proceso de captura y almacenamiento se divide en tres fases las cuales se indican a continuación:

- Capturar o secuestrar: capturar o secuestrar por medio de los árboles o sombra que se encuentra en el sistema agroforestal el CO<sub>2</sub> emitido en el proceso de la cosecha, separándolo así de otros gases que pudieran estar presentes en este proceso.
- Transportar o movilizar: En la etapa de transporte, el CO<sub>2</sub> capturado es dirigido hacia los denominados sumideros de carbono.
- Almacenamiento: el almacenamiento del CO<sub>2</sub> es la etapa donde por medio de un proceso físico o biológico, el gas previamente secuestrado y transportado es almacenado en los sumideros, los cuales pueden ser el bosque o el suelo directamente.

Gracias a este proceso se busca evitar que este GEI se emita y acumule en la atmósfera.

Figura 8. **Sumideros de carbono**



Fuente: Comisión Nacional de Áreas Protegidas (CONAP). (2015). *¿Qué son los sumideros de carbono?* Consultado el 15 de octubre de 2021. Recuperado de [https://twitter.com/semarnat\\_mx/status/674457246491914241](https://twitter.com/semarnat_mx/status/674457246491914241).

### 7.3.6. Datos, información y conocimiento

A continuación, en los siguientes incisos se describen los datos, información y conocimientos necesarios para esta investigación.

### **7.3.6.1. Datos**

Desde siempre, en cada proceso o actividad que se desarrolla a lo largo de la vida cotidiana, por ejemplo, actividades realizadas en el hogar, en el gimnasio, en el trabajo, la escuela y muchas otras más, se está ante la presencia de grandes cantidades de datos, donde estos pueden convertirse en fuente de información importante al momento de realizar algún estudio o investigación.

Se puede decir que los datos (datos en bruto) no brindan una perspectiva de la situación que se está analizando, y que por sí solos al momento de llevar a cabo una toma de decisiones en cualquier ámbito son irrelevantes, el año de nacimiento o la altura de una persona por si solos son simples números que carecen de utilidad y no dan visual a ninguna situación en específico. Los datos pueden ser de tipo cualitativo o cuantitativo, como por ejemplo la edad de una persona, el color de ojos de una persona, el nombre o la altura de la misma y pueden ser almacenados en diversos lugares, como por ejemplo un papel, un archivo plano, un dispositivo extraíble e inclusive dentro de la mente de un individuo.

### **7.3.6.2. Información**

El termino información se le da al momento en que el conjunto de datos ha sido parte de un proceso y que es relevante para el analista u observador de estos, es decir, un conjunto de datos se convierte en información al momento en que estos tienen 3 factores clave, siendo estos:

- Relevancia
- Propósito
- Contexto.

La información a diferencia de los datos crudos, brindan una perspectiva más clara de la situación de las cosas, permitiendo realizar análisis más profundos, y capacidad de toma de decisiones en función de lo que se desea.

Para poder llevar a cabo la transformación de datos en crudo hacia la información como tal, se debe agregar algún tipo de valor a los mismos, dentro de este se puede hacer mención a las siguientes cinco variables (a las que se denominarán como las 5C) que le agregan valor a un dato:

- Contexto: Ayuda a determinar el propósito con el cual o para que se utilizara el dato.
- Categoría: Permite clasificar los datos según la categoría o unidad de medida que le interese al investigador.
- Cálculo: Ya sea proceso de tipo matemático o estadístico.
- Corrección: Este valor se agrega al realizar una revisión de los datos y generar un filtrado de estos para eliminar todos aquellos que no son útiles.
- Condensación: Similar a la categoría de datos, pero en este caso los datos se han logrado agrupar o resumir de una forma más compacta para su estudio.

### **7.3.6.3. Conocimiento**

El termino conocimiento se puede conceptualizar como el valor añadido a las experiencias vividas, en conjunto con la información que se ha adquirido a lo largo de estas experiencias, gracias a esto sirve como base para adquirir nuevas experiencias e información.

El conocimiento, corresponde al escalón más alto dentro de los datos y de la información, si se ejemplifica desde un esquema piramidal se verían de la siguiente forma:

Figura 9. Pirámide de conocimiento



Fuente: Sinnexus. (2021). *Datos, información, conocimiento*. Consultado el 15 de septiembre de 2021. Recuperado de [https://www.sinnexus.com/business\\_intelligence/piramide\\_negocio.aspx](https://www.sinnexus.com/business_intelligence/piramide_negocio.aspx).

Para poder realizar la transformación de la información y que esta sea considerada como conocimiento, debe cumplir con diversas acciones que se enumeran a continuación:

- Poder ser comparada con otros elementos (tener punto de comparación con otros elementos).
- Poder de inferir o predecir sobre acciones futuras.
- Búsqueda e identificación de patrones que permitan visualizar conexiones.
- Poder de entablar una conversación e intercambiar conocimiento con otros similares.

### 7.3.7. Indicadores

Es importante el poder tener un estándar de medición con el cual se realicen comparaciones entre datos a lo largo del tiempo, y es acá donde surge

un concepto muy importante denominado Indicador. Se puede decir que un indicador es una herramienta de tipo estadístico o matemático que permite tener el manejo, control y visualización de los cambios que han sufrido los datos a lo largo de un periodo de tiempo establecido.

Los indicadores son ampliamente utilizados en diversas áreas industriales para medir el rendimiento de diferentes actividades o procesos, por ejemplo:

- En una empresa cuyo giro del negocio es el de manufacturación y venta de productos es posible observar dos indicadores indispensables como lo son el indicador de venta mensual y el indicador de calidad de la producción.
- En actividades cotidianas como el deporte también se pueden usar indicadores, en este caso utilizados para determinar el rendimiento que tiene el deportista cuando se somete a condiciones especiales como clima, topografía, uso de nuevas prendas deportivas, entre otros.

Existen diversos tipos de indicadores como, por ejemplo:

- Rentabilidad
- Productividad
- Innovación
- Desempeño
- Entre otros

En la actualidad, gracias a la revolución industrial denominada como 4.0 se ha dado grandes cambios en la forma en que se llevan a cabo muchos procesos y actividades en nuestra vida cotidiana, y el uso de datos como fuente de información y los indicadores como herramienta de medición han tomado un

papel sumamente importante en todo el mundo ya que los nuevos productos y servicios que se brindan al cliente están basados en toda esta información recolectada.

### **7.3.8. Estadística**

Es una división derivada de la matemática, esta se encarga del estudio y análisis de muchas actividades tanto dentro de la industria como dentro de lo más cotidiano de nuestra vida. La estadística es una excelente y poderosa herramienta, que con el conocimiento suficiente sobre el uso de sus diversas técnicas y donde aplicar cada una de ellas, permite al usuario realizar análisis de datos confiables en diversos campos de estudio.

#### **7.3.8.1. Medidas de tendencia central**

Una medida de tendencia central dentro de estadística se ve representada como una variable de estudio que corresponde a un grupo de datos, los cuales se encontraran distribuidos por todo el eje X (eje horizontal) de un gráfico de datos. A continuación, una breve descripción de estas variables:

- Media

La media también conocida como media aritmética o promedio es una de las medidas de tendencia central, y muestra el punto exacto donde la variable a analizar alcanza el mayor número de datos acumulados, y se describe como la media de un conjunto de  $n$  mediciones es igual a la suma de las mediciones dividido  $n$ .

La media poblacional se ve representada por el símbolo  $\mu$  (mu) mientras que la media de la muestra se ve representada por el símbolo  $\bar{x}$  (x barra); en forma de ecuación, la media se ve representada de la siguiente forma:

$$\sum_{i=1}^n X_i \text{ que significa } X_1 + X_2 + X_3 + \dots + X_n$$

(Ecuación 1)

Donde:

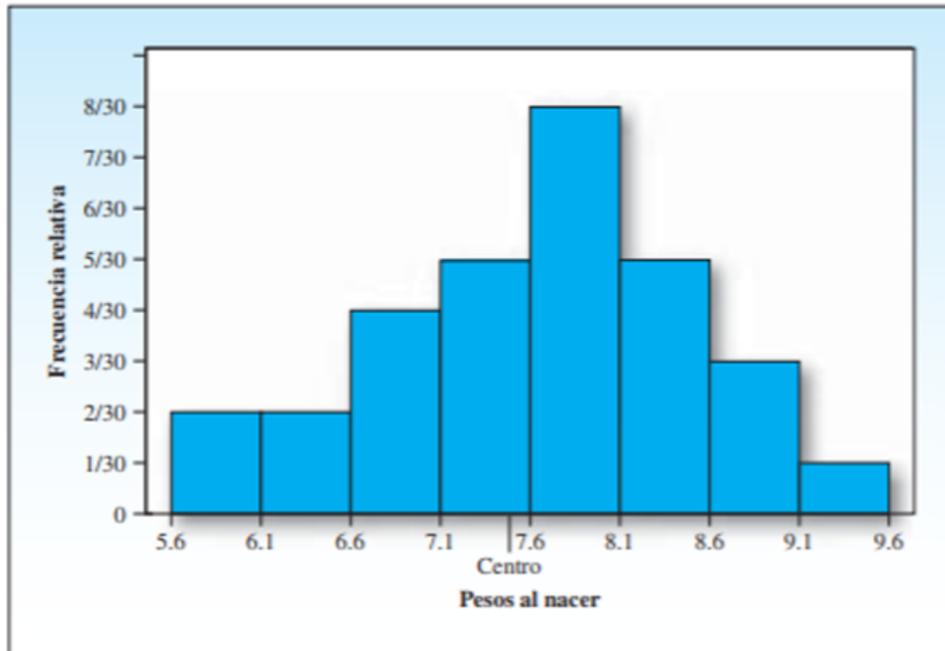
$i$  = índice del dato

$n$  = total de datos

$X_i$  = dato correspondiente a ese índice

La representación gráfica de esta medida se puede representar como en la siguiente gráfica:

Figura 10. Representación gráfica de la media



Fuente: Mendenhall, Beaver y Beaver. (2010). *Introducción a la probabilidad y estadística*.

- Mediana

La mediana también pertenece al conjunto de medidas de tendencia central, en la cual luego de un proceso de ordenamiento de datos ascendente (menor a mayor), se logra determinar el dato ubicado justo al medio de un conjunto de  $n$  datos.

En dado caso existan dos datos que representen el punto medio de un grupo de datos, se debe realizar un promedio entre ambos para determinar un único punto medio.

$$m = \frac{\text{dato 1} + \text{dato 2}}{2}$$

Donde:

m = mediana

dato 1 = dato medio inferior

dato 2 = dato medio superior

La mediana se ve representada por la letra m (m) tanto para muestra como para la población. Cabe resaltar que cuando una distribución de datos es simétrica, tanto la media como la mediana son iguales.

- Moda

La moda es utilizada generalmente para el análisis de grandes cantidades de datos, e indica el dato o categoría que se presenta con mayor frecuencia dentro de un grupo de datos de tamaño n.

Para poder seleccionar la moda dentro de un grupo de datos crudos, es necesario llevar a cabo un proceso para el ordenamiento y clasificación de datos y con base en este poder realizar la creación de una tabla de frecuencias de tipo histograma, con la cual se puede observar con mayor facilidad el pico más alto del histograma, el cual se denomina como clase modal correspondiente a un conjunto de datos y a la vez, el punto medio de esta clase modal se denominara como la moda de este conjunto de datos.

Dentro de una distribución y clasificación de datos puede existir una o más modas, y a diferencia de la mediana, se puede encontrar más de una moda en el grupo de datos estudiado.

Por lo cual se puede concluir, entonces, que la moda indica el dato o datos que se repite mayor cantidad de veces dentro de un conjunto clasificado de datos.

Tabla III. **Ejemplo de moda estadística**

<b>Edades (años)</b>	<b>Frecuencia</b>
15-20	10
21-26	5
27-32	7
33-38	12
39-44	8

Fuente: elaboración propia, empleando Microsoft Excel 365.

Como se observa (Tabla III. Ejemplo de moda estadística) se tiene la clasificación de edades para un grupo de estudio con su respectiva frecuencia de datos, el grupo de entre 33 y 38 años tiene la mayor cantidad de datos, por lo cual, se puede decir entonces que la moda del grupo de datos se ubica en el rango de entre 33 y 38 años con una frecuencia de 12 muestras.

#### **7.3.8.2. Medidas de dispersión**

También conocidas como medidas de variabilidad, son herramientas muy útiles que permiten identificar o cuantificar que tanta variación o distancia existe entre los datos que corresponden a una muestra o población.

Regularmente se utilizan luego del análisis de las medidas de tendencia central para verificar y validar la integridad de los cálculos y los resultados que fueron obtenidos, ya que Con base en estos es posible realizar la toma de decisiones.

La confiabilidad de los datos versus la medida de dispersión es inversamente proporcional, es decir, a mayor medida de dispersión menor confianza en los datos, por lo cual es posible señalar lo siguiente: A razón de que una medida de dispersión se aleje de cero (0), las medidas o datos de tendencia central indicados anteriormente son cada vez menos representativos y por lo cual se debe analizar si los datos capturados en la muestra son confiables para la toma de decisiones.

Las tres medidas de dispersión más utilizadas a nivel estadístico son: El Rango, la varianza y la desviación estándar.

- Rango

El rango, el cual es identificado con el símbolo R (R), se conoce como la medida que indica la diferencia entre el valor máximo o más grande de un grupo de datos respecto al valor mínimo o más pequeño del mismo grupo de datos.

Entre más grande sea el rango del grupo de datos, mayor será la dispersión de estos.

$$R = \text{dato mayor} - \text{dato menor}$$

Donde:

R = rango

Dato mayor = dato más grande de la serie de datos

Dato menor = dato más pequeño de la serie de datos

Para llevar a cabo el cálculo del rango en un conjunto de datos, al igual que con la moda es necesario realizar un ordenamiento de datos previo,

organizándolos del menor al mayor dato y luego aplicando la formula previamente anotada.

- Varianza

La varianza, la cual se identifica con el símbolo  $\sigma^2$  si es para población y con el símbolo  $s^2$  si es para la muestra se define como la media de la diferencia o distancia entre los datos por individual de un conjunto de datos respecto al valor de la media obtenida del mismo conjunto de datos elevada al cuadrado.

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}, s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

Donde:

$\sigma^2, s^2 =$  varianza

$x_i =$  dato

$\mu, \bar{x} =$  media

$N, n =$  total de los datos

- Desviación estándar

Corresponde a otra medida de dispersión que se identifica con el símbolo  $\sigma$  si es para población y con el símbolo  $s$  si es para la muestra, y representa la distancia o separación que existe entre los datos correspondientes a un mismo grupo tabulado de datos.

$$\sigma = \sqrt{\sigma^2}, s = \sqrt{s^2}$$

Donde:

$\sigma^2, s^2 =$  varianza

### **7.3.9. Análisis de datos**

Es sumamente importante y una gran herramienta utilizada actualmente por toda la industria, ya que brinda información relevante y precisa sobre los gustos, requerimientos, expectativas, entre otras de los clientes; Con base en un análisis de datos es posible realizar grandes acciones, como por ejemplo identificar el mercado objetivo y los gustos de estos para:

- Innovar en un nuevo producto o servicio
- Innovar en un producto o servicio existente
- Crear un nuevo proyecto o empresa

### **7.3.10. Bases de datos**

Durante el transcurso de los años se ha utilizado diversos medios o dispositivos físicos para el almacenamiento y manipulación de la información. Actualmente las bases de datos corresponden a espacios donde se almacena información de forma ordenada o estructurada sistemáticamente para acceder a ella posteriormente y poder gestionarla de la mejor forma posible.

El uso y manejo de una base de datos cuando es realizado por medio de un sistema gestor de base de datos brinda diferentes ventajas, dentro de estas se puede mencionar:

- Consultas varias
- Flexibilidad

- Integridad de los datos
- Seguridad
- Entre otros

Existen una gran variedad de bases de datos, y su uso depende de las expectativas o la necesidad del usuario, dentro de estas se puede mencionar:

- Bases de datos dinámicas o estáticas
- Bases de datos bibliográficas
- Bases de datos jerárquicas
- Bases de datos relacionales
- Bases de datos no relacionales
- Bases de datos documentales
- Entre otras

Cada una de las anteriormente mencionadas tiene diversas características, ventajas y desventajas según el uso para el cual se desee implementar. Actualmente las bases de datos se están actualizando de tal forma que están migrando o moviéndose a un sistema de almacenamiento denominado como la nube.

La nube, es un nuevo servicio de base de datos en internet donde los usuarios pagan por un espacio virtual (para el usuario) y pueden almacenar su información, para luego poder acceder a ella desde cualquier dispositivo digital en cualquier parte del mundo.

Sin embargo, el concepto de espacio virtual únicamente aplica para el usuario que contrata el servicio, ya que las empresas que brindan el servicio

cuentan con los espacios físicos y tangibles donde almacenan toda la información de los clientes.

### **7.3.11. Técnicas de análisis de datos**

Las herramientas estadísticas y la estadística tal cual, ha sido utilizada como la técnica de análisis de datos más accesible y confiable en el manejo de datos, y al analizar información, sin embargo, actualmente, gracias a la evolución de la tecnología y las técnicas de información se han generado nuevas herramientas más especializadas y con mayor eficiencia al momento de manipular información.

Dentro de las técnicas utilizadas para analizar datos se puede mencionar:

- Árboles de decisión
- Aprendizaje automático y
- Minería de datos

### **7.3.12. Árboles de decisión**

Los árboles de decisión constituyen una metodología en función del aprendizaje y son herramientas que se encargan de clasificar o segmentar la información en función de una serie de reglas previamente establecidas por el analista, la clasificación de esta se ve representada de forma visual en forma de árbol, de ahí su nombre representativo.

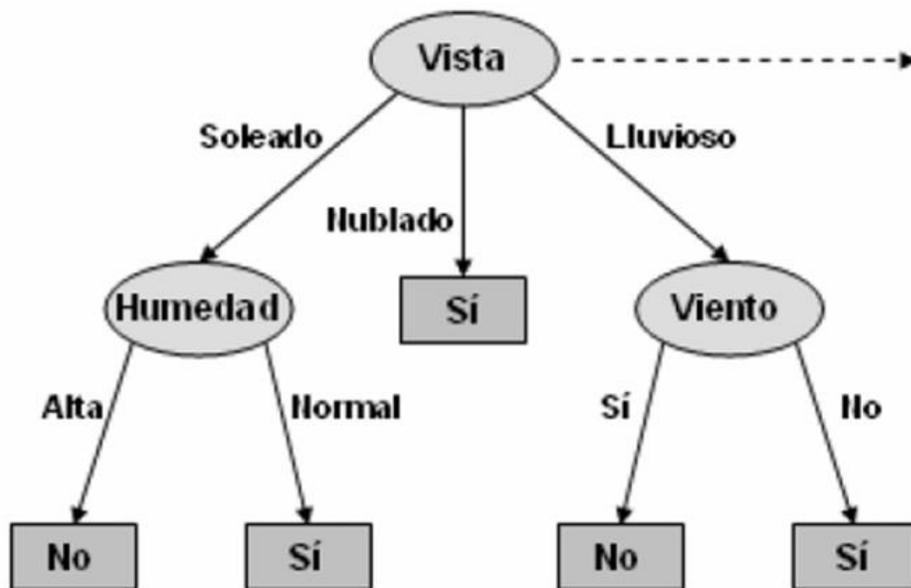
La información se ve reflejada como un vector de pares ordenados (atributo – valor) la cual será la etiqueta para cada eje del árbol cada clase

representa las hojas de este; el árbol va creciendo desde su raíz y la trayectoria de las hojas se determinan en función de las reglas establecidas previamente.

La mayoría de los sistemas de aprendizaje basados en técnica de árboles regularmente no son de tipo incremental, existen diversos algoritmos de árboles, sin embargo, a continuación, se listará a tres (3) de estos:

- Sistema ID3: el árbol de sistema ID3 es un algoritmo simple, en el cual se debe seleccionar un atributo o variable como raíz del árbol y cada rama se creará en función de las reglas establecidas para cada valor que pueda tomar esta variable. Este proceso se repetirá de forma cíclica hasta obtener la clasificación final de los datos según la variable elegida como raíz.

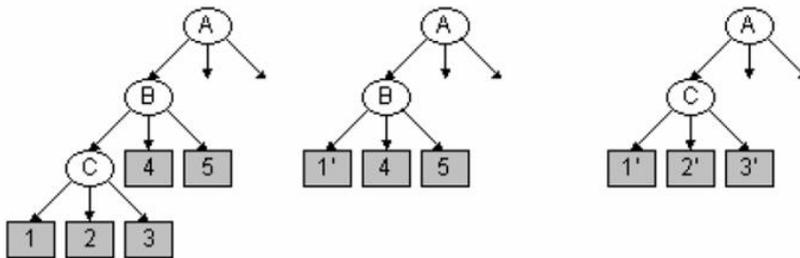
Figura 11. **Árbol tipo ID3**



Fuente: Molina y García. (2006). *Técnicas de análisis de datos*.

- Sistema C4.5: este árbol tiene la capacidad de elaborarse con atributos de valores tanto discretos como continuos, en dado caso la información a analizar es de tipo discreta, el árbol tendrá tantas ramas como la cantidad de valores que pueda llegar a tomar el atributo; si la información es de tipo continua, el árbol no clasificara correctamente la información, sin embargo, existe una serie de pasos que pueden corregir esta situación.

Figura 12. **Árbol tipo C4.5**



Fuente: Molina y García. (2006). *Técnicas de análisis de datos*.

- **Árbol de un solo nivel:** el árbol de un solo nivel es el más básico de los tres, sin embargo, en función de la información y la necesidad del uso de esta, puede llegar a brindar resultados interesantes, este árbol no tiene opciones de configuración y únicamente muestra tres ramas, donde en una de ellas muestra atributos desconocidos, las restantes clasifican la información en función del valor del atributo previamente establecido.

### 7.3.12.1. **Aprendizaje automático**

El aprendizaje automático forma parte de la vida cotidiana de un individuo, tanto en actividades deportivas, laborales, personales, u otras, una persona puede tomar decisiones a futuro en función de experiencias vividas previamente.

Tomando en cuenta este concepto, cuando se habla de información se puede decir que el aprendizaje automático es una herramienta tecnológica que consiste en el aprendizaje de conductas, patrones o reglas a partir de la información o la experiencia.

Existen diversas técnicas que se basan en la lógica del aprendizaje automático como, por ejemplo: el aprendizaje conceptual, redes neuronales, arboles de decisión, entre otros. Con base en las distintas investigaciones realizadas a partir de los resultados resultantes del aprendizaje automático, se han creado tres (3) categorías siendo estas:

- Aprendizaje activo: se lleva a cabo el aprendizaje y modificación de comportamientos de forma paralela a la vez que se analiza la información en tiempo real.
- Aprendizaje desde el conocimiento: se modifica el comportamiento del modelo basándose en la información obtenida de sucesos o eventos previos.
- Aprendizaje incremental: conocido también como inteligencia artificial, este es el aprendizaje más complejo de tipo continuo y tiene como base tanto el aprendizaje activo como el aprendizaje desde el conocimiento, siempre está en constante cambio.

### **7.3.12.2. Minería de datos**

La minería de datos es una herramienta que ha sido utilizada desde antes de darle este nombre, y se compone de: resultados, técnicas y a la vez otras herramientas que son utilizadas para la recopilación de datos transformados en información útil de bases de datos extremadamente amplias.

Esta técnica denominada como minería de datos es utilizada en diversos campos de estudio como, por ejemplo:

- Estadísticos
- Analistas de datos
- Administradores de sistemas informáticos

La recopilación y análisis de la información puede llevarse a cabo por medio de métodos cotidianos o manuales (aunque tomaría un tiempo extremadamente amplio el cual a la vez incurriría en un aumento en los costos administrativos), sin embargo, cuando la base de datos o la cantidad de información sobrepasa la capacidad de análisis, es momento de acudir a una nueva técnica y es aquí donde se usa la minería de datos.

Existen diversos conceptos teóricos para esta gran herramienta, pero para el actual trabajo se hará referencia a la técnica de minería de datos como: el procedimiento de recopilar información relevante a un tema en específico, proveniente de grandes almacenes de datos y que permite observar, analizar y determinar tendencias o patrones entre los mismos.

Para poder realizar un adecuado procedimiento de minería de datos, es necesario tomar en cuenta algunas herramientas alternativas para que los datos sean analizados correctamente, entre ellas, se menciona:

- Buen razonamiento estadístico
- Técnicas de visualización
- Procesamiento paralelo
- Sistemas de apoyo
- Aprendizaje automático

Con base en lo planteado anteriormente se puede indicar diversas áreas de aplicación para la minería de datos como, por ejemplo:

- *Marketing*
- Banca y compañías de seguros
- Telecomunicaciones
- Clima
- Industria farmacéutica
- Biología
- Agronomía
- Entre otros

Para realizar un buen análisis de datos al momento de estar minando datos es básico realizar cuatro (4) pasos esenciales:

- Determinar el objetivo general: el analista debe establecer tanto las variables determinantes, como el objetivo principal que desea cumplir previo a realizar el análisis y minería de datos.
- Manipulación o procesamiento de datos: efectuar la correcta manipulación de datos que fueron recopilados; realizar la selección de la información relevante, realizar la limpieza para la reducción y transformación de los datos, para brindarles un valor agregado y utilizar información útil para generar la nueva base de datos.
- Determinar el modelo de datos: con base en el análisis estadístico realizado previamente, y posteriormente la visualización de la información por medio de herramientas graficas se puede conocer si hay relación alguna entre la información para identificar un modelo de datos adecuado.
- Análisis e interpretación de los resultados: Al momento de contar con el modelo de datos seleccionado, se debe proceder con la verificación del

modelo y analizar los resultados obtenidos por medio de este, para determinar si son coherentes y coinciden con lo esperado al inicio del estudio.

### **7.3.13. Software de análisis estadístico**

El manejo y análisis correspondiente de la información, es fundamental al momento de tomar decisiones, ya que, hoy en día la información es la clave principal para una buena gestión en cualquier actividad ya sea de tipo personal o industrial.

Por eso, tener conocimiento sobre herramientas que permitan aplicar todos los conocimientos sobre el análisis de datos es tan importante en la actualidad.

Estas herramientas conocidas como software de análisis estadísticos permiten llevar a cabo más eficiente y confiablemente el manejo y análisis de grandes volúmenes de información.

Existen diversas herramientas para el análisis de datos como, por ejemplo:

- Rstudio
- Python
- Excel
- Tableau
- Power Bi
- Entre otros

### 7.3.13.1. Rstudio

El software de análisis de datos conocido como Rstudio se basa en el lenguaje de programación R, y utiliza un entorno para la programación y creación de gráficos estadísticos, actualmente es una de las mejores herramientas de tipo analítico, donde se puede llevar a cabo diversas acciones para la manipulación de información.

Principalmente Rstudio se utiliza para la creación de modelos de datos y manipulación estadística de los mismos, es muy fácil e intuitiva de utilizar, sin embargo, es necesario poseer conocimiento técnico de programación para llevar a cabo el uso de este.

Rstudio brinda al usuario un extenso catálogo de utilidades o herramientas estadísticas, dentro de las cuales se puede mencionar:

- Modelado lineal
- Pruebas estadísticas clásicas
- Series de tiempos
- Clasificación de datos
- Agrupamiento de datos
- Técnicas de presentación graficas
- Entre otros

Rstudio es de uso libre y se puede ejecutar en diversos sistemas operativos y plataformas, incluyendo dentro de ellas las más utilizadas en el mercado (Windows, Linux y MacOS).

Rstudio consta de 4 espacios básicos para el usuario, siendo estos: editor de código, consola, directorio de trabajo y por último los *plots* y paquetes.

- Editor de código: esta área es el espacio disponible para que el usuario escriba el código fuente para los programas o análisis estadísticos deseados y pueda posteriormente almacenarlos.
- Consola: la consola es la pantalla de salida de Rstudio, es acá donde se muestra al usuario los resultados de las líneas de código que se ejecutan en el editor de código.
- Directorio de trabajo: en el directorio de trabajo se muestra todo el historial de variables que se ha ido creando en cada programa o análisis realizado.
- *Plots* y paquetes: el área de *plots* y paquetes está destinado para visualizar los paquetes secundarios que se han instalado como herramienta para el análisis de datos.

## 8. PROPUESTA DE ÍNDICE DE CONTENIDOS

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES

LISTA DE SÍMBOLOS

GLOSARIO

RESUMEN

PLANTEAMIENTO DEL PROBLEMA

OBJETIVOS

RESUMEN MARCO TEÓRICO

INTRODUCCIÓN

1. ANTECEDENTES

2. JUSTIFICACIÓN

3. MARCO TEÓRICO

4. MODELO DE DATOS

4.1. Requerimientos

4.2. Manipulación de los datos

4.2.1. Filtrado de datos

4.2.2. Clasificación de datos

4.3. Desarrollo de modelo de datos

4.3.1. Alimentación de base de datos

4.3.2. Clasificación de datos para entrenamiento y evaluación

4.3.3. Creación del modelo de datos

- 4.4. Entrenamiento del modelo de datos
  - 4.4.1. Pruebas técnicas
- 4.5. Evaluación del modelo de datos
  - 4.5.1. Indicadores por evaluar
  - 4.5.2. Pruebas técnicas
- 5. PRESENTACIÓN Y DISCUSIÓN DE RESULTADOS
  - 5.1. Diseño de la solución
  - 6.2. Comparación con modelos similares
  - 6.3. Gráficos de datos

CONCLUSIONES

RECOMENDACIONES

REFERENCIAS BIBLIOGRÁFICAS

ANEXOS

## **9. METODOLOGÍA**

### **9.1. Características del estudio**

El estudio que a continuación se presenta, se basará en un enfoque de tipo cualitativo ya que el objetivo del trabajo es lograr cuantificar la huella de carbón producida durante la cadena de producción del SAF café y a la vez conocer o determinar cómo se ven relacionadas las distintas variables y su porcentaje de participación dentro de esta emisión.

Este se desarrollará con base en la data obtenida por medio de las mediciones realizadas por la Asociación Nacional del Café –Anacafé-.

Para ello, el alcance del proyecto será de tipo descriptivo correlacional, ya que, con base en los datos brindados por Anacafé, se llevará a cabo un análisis de los mismos y se tendrá como objetivo el determinar la cantidad de huella de carbón que interviene (porcentaje) en el proceso de producción del SAF café, así también se podrá determinar el (índice/porcentaje) en que afecta cada una de las variables de estudio involucradas en el proceso respecto a la emisión de dióxido de carbono en sus respectivas etapas.

Respecto al diseño del proyecto, se llevará a cabo en un modelo de diseño no experimental ya que para desarrollar la formulación de modelo de datos para estimar la huella de carbono en la cadena de producción del sector SAF café en Guatemala no habrá manipulación previa o posterior de los datos de estudio brindados por Anacafé, en función de lo previamente mencionado se determina que además será un diseño de forma transversal descriptivo, pues el análisis de

los datos se llevara a cabo para un determinado tiempo en específico en el que se desee conocer la cantidad de huella de carbono generada en ese preciso momento, es decir, no habrá un estudio de tiempo secuencial para los datos.

## 9.2. Unidades de análisis

Este trabajo se enfocará en el sector cafetalero de Guatemala que conforma a la Asociación Nacional del Café –Anacafé- como población de estudio, la cual se subdivide en ocho (8) diferentes regiones a lo largo del territorio nacional, y de las cuales se llevara a cabo un muestreo estratificado para específicamente 3 regiones, las cuales serán el grupo de muestreo para el proyecto. (Se puede colocar las 3 regiones en mención y algunas características de estas para dar mayor visual al lector).

## 9.3. Variables

Las variables de estudio se describen a continuación:

Tabla IV. **Variables de estudio**

<b>Variable de estudio</b>	<b>Definición teórica</b>	<b>Definición operativa</b>
Residuo de pulpa	Es todo aquel residuo restante de tipo biodegradable que proviene de diversas etapas durante el proceso productivo del SAF café, se desintegra de forma rápida convirtiéndose en otro tipo de materia orgánica.	Durante el proceso de producción del SAF café se genera residuos, provenientes de distintas actividades como lo son el sistema agroforestal, benéfico húmedo, beneficio seco y Trillado.  Equivalente de CO <sub>2</sub> generado por la gestión de la pulpa durante el proceso productivo del SAF café. (p.e.CO <sub>2</sub> /qq)

Continuación de la tabla IV.

Combustible	Recurso natural líquido con el cual se lleva a cabo distintas actividades dentro del proceso de producción del SAF café, dentro de ellas se puede mencionar: el despulpado, lavado, transporte, entre otros.	El combustible puede ser utilizado en distintos procesos dentro de la producción del SAF café, tanto dentro del sistema agroforestal como en el beneficiado húmedo y trillado. Equivalente de CO <sub>2</sub> generado por uso de combustibles.
Sombra	Característica del sistema agroforestal que se encuentra distribuida en una determinada región o zona y que ayuda a satisfacer diversas necesidades de su entorno.	La cantidad de sombra que se encuentre distribuida a lo largo del sistema agroforestal contribuirá con la captación de CO <sub>2</sub> que se genere dentro del mismo.  Cantidad de CO <sub>2</sub> captado por sombra.
Energía eléctrica	Fuente de energía renovable que puede obtenerse por diversos medios y sirve como combustible o fuente de energía para la maquinaria, medio de transporte, entre otros.	La energía eléctrica se ve involucrada en la mayor parte de etapas del proceso de producción del SAF café, tomando así gran relevancia en la generación de CO <sub>2</sub> dentro del proceso.  Cantidad de CO <sub>2</sub> generado por KW/mes consumido.
Insumos agrícolas	Productos de uso agropecuario utilizados al inicio y durante el cultivo con el objetivo de tener una buena cosecha. Existen productos agrícolas mayormente utilizados, tales como semillas, fertilizantes y productos para el control de plagas y enfermedades.	Los insumos agrícolas se aplican al cultivo del SAF café durante la fase del sistema agroforestal, dentro de ellos se puede encontrar los abonos, pesticidas, entre otros, que toman relevancia en la generación de CO <sub>2</sub> dentro del proceso de producción del SAF café.  Cantidad de CO <sub>2</sub> generado por insumos agrícolas consumidos
Huella de carbono	Variable de medición que es utilizada para cuantificar el impacto que una actividad tiene sobre el cambio climático.	Variable por estudiar y cuantificar dentro del sistema de producción del SAF café.

Continuación de la tabla IV.

Altitud	La altura es la unidad de medida utilizada para definir la distancia vertical entre un punto determinado en el espacio respecto a otro, en este caso ese punto en el espacio se mide respecto a la superficie del mar.	La variable altitud se analizará para validar si existe alguna relación entre la altura a la cual se encuentra la unidad productiva y la cantidad de CO <sub>2</sub> emitido
Temperatura	La temperatura es la unidad de medida utilizada para expresar la sensación térmica de un lugar u objeto, puede medirse por medio de un termómetro y permite tener noción de clima frío o caliente.	La variable temperatura se analizará para validar si existe alguna relación entre la temperatura a la cual se encuentra sometida la unidad productiva y la cantidad de CO <sub>2</sub> emitido

Fuente: elaboración propia, empleando Microsoft Excel 365.

#### **9.4. Fases del estudio**

A continuación, se describen las fases del estudio.

##### **9.4.1. Revisión documental**

Durante la revisión documental se recurrirá a la revisión de diversas fuentes de datos que contengan información relevante a los distintos temas involucrados dentro de este trabajo de graduación.

Se revisará documentación relacionada a temas de:

- Café
- Sistemas agroforestales
- Producción de café
- Software de datos
- Herramientas de datos

- Programación
- *Machine Learning*
- Normas y leyes medio ambientales
- Entre otros

#### **9.4.2. Recepción de información (datos crudos)**

La información o datos crudos con los cuales se procederá a trabajar es la base fundamental en este estudio, ya que sin ellos no podrá llevarse a cabo ningún análisis estadístico y por lo cual no se podrá proceder a realizar ningún modelo matemático que permita contabilizar la huella de carbón en los sistemas productivos del SAF café.

Para este estudio en específico, no se llevará a cabo la recopilación de datos desde cero, ya que los datos con los cuales se trabajará serán proporcionados por Anacafé con base al histórico con que cuentan en sus bases de datos.

#### **9.4.3. Filtrado de información**

El filtrado de información es el proceso por el cual, una vez obtenidos los datos crudos, estos son sometidos a una serie de pasos donde se seleccionan los datos más relevantes que sirvan como función de carácter importante y relevante a la investigación.

En el presente estudio el filtrado de información se llevará a cabo para las tres (3) regiones indicadas en el documento, tomando en consideración las distintas variables:

- Energía eléctrica
- Combustibles
- Huella de carbón
- Altitud
- Entre otras

Se determinará cuáles datos se encuentran fuera de los rangos deseados para el estudio por medio de gráficos de control y otras herramientas para determinar los datos que causaran posible ruido al momento de trabajar el modelo, gracias a este filtrado de información se limpiará la data y se podrá a dar inicio al análisis de estos.

#### **9.4.4. Análisis de información por medio de métodos estadísticos**

Una vez los datos han sido filtrados y se ha determinado el grupo de datos útiles para llevar a cabo el estudio, se procederá a utilizar los diferentes métodos estadísticos para analizar los diversos comportamientos que se dan en los grupos de datos, se llevará a cabo la ilustración de estos comportamientos por medio de gráficos.

#### **9.4.5. Clasificación de la información**

Una vez la data ha sido limpiada y analizada, llego el momento de determinar su función dentro del modelo de datos y proceder a la clasificación de estos. Los datos se dividirán en dos grupos siendo estos:

- Grupo de datos para el entrenamiento del modelo principal
- Grupo de datos para la evaluación del modelo principal

Por medio de software como R se determinará la cantidad adecuada u optima que debe realizarse para cada una de las clasificaciones.

#### **9.4.6. Desarrollo del modelo de datos**

En la fase del desarrollo del modelo de datos, se contemplarán todas las actividades realizadas previamente y se buscara por medio de métodos matemáticos y el uso de software especializado el determinar la correlación que existe por cada una de las variables con la cantidad de huella de carbono generada en el sistema de producción del SAF café. Sera importante determina este indicador ya que, con base en él se podrá realizar los ajustes al modelo de datos.

#### **9.4.7. Entrenamiento del modelo de datos**

Al momento de contar con el modelo de datos desarrollado, se procederá al entrenamiento de este, esto utilizando el grupo de datos particionado en el proceso de la clasificación de la información, para el entrenamiento de datos se utilizará el método de árbol con el cual se buscará tener el menor error de dispersión posible.

#### **9.4.8. Evaluación del modelo de datos**

Al igual que con el entrenamiento del modelo de datos, es momento de proceder con la evaluación final del mismo, para esto se usará el grupo de datos o data set particionado específicamente para este proceso, estos datos permitirán evaluar el correcto funcionamiento del modelo y se cotejarán con los resultados obtenidos en la fase del entrenamiento.

El error de dispersión puede variar respecto al obtenido en el entrenamiento del modelo, pero debe encontrarse dentro de un rango aceptable para el analista.

#### **9.4.9. Entrega final del modelo de datos**

Al momento de contar con las pruebas finales y tener total confianza de que el modelo de datos es funcional, llega el momento de presentar el producto final a la Asociación Nacional del Café –Anacafé-.

## **10. TÉCNICAS DE ANÁLISIS DE INFORMACIÓN**

A continuación, se describe brevemente algunas de las técnicas o herramientas de carácter estadístico descriptivo inferencial que se utilizarán durante el desarrollo del modelo de datos correspondiente a este trabajo de investigación.

### **10.1. Preparación de la información**

Para poder proceder al análisis de la información y a la creación del modelo de datos, es imprescindible preparar los datos de estudio que Anacafé brindo como apoyo al trabajo de investigación para el respectivo análisis, por lo cual, para esta etapa del proyecto se deberá:

#### **10.1.1. Creación de la base de datos plana**

Por medio de Excel se llevara a cabo una base de datos plana, en la cual se almacenará los datos brindados por Anacafé, de esta forma se tendrá acceso a los datos de forma rápida y organizada por medio de un archivo con extensión .CSV el cual podrá ser visualizado más adelante en el software de análisis estadístico RStudio.

#### **10.1.2. Carga de información al software de análisis**

Al momento de tener preparada la base de datos plana, es momento de cargar la información al software que se utilizara para el análisis de la misma.

Para llevar a cabo este modelo se utilizará el lenguaje de programación R por medio del software RStudio.

### **10.1.3. Clasificación de la segmentación de la muestra**

Se agregará una nueva columna denominada segmento productivo la cual identificará a cada unidad productiva según su capacidad de producción dentro del sector SAF café, esta clasificación de se realizará con el objetivo de poder visualizar de una forma más detallada los resultados posteriormente, así también se podrá observar resultados en función de la segmentación de la población.

### **10.1.4. Selección de tamaño de la muestra**

Previo a iniciar el análisis de los datos es importante determinar el tamaño de la muestra necesaria para la clasificación de datos que se utilizaran con el objetivo de entrenar el modelo de datos y el tamaño de la muestra que se utilizara para llevar a cabo la validación del modelo de datos. Es importante conocer los parámetros que serán la base para determinar el tamaño de la muestra, es decir, identificar si se realizara por medio de la clasificación de la segmentación de la muestra o por otra variable relevante para el investigador, por lo cual es importante hacerse la pregunta ¿Qué variable o variables serán de impacto para determinar el tamaño de la muestra?

Al momento de conocer la respuesta a este cuestionamiento es momento de determinar el tamaño de la muestra por medio de dos métodos:

- Por medio del uso de estadística descriptiva utilizando el método de tamaño de la muestra.
- Por medio de la herramienta de árboles de decisión.

Con lo cual se realizará dos grupos de datos clasificándose como:

- Unidades productivas para entrenamiento de modelo
- Unidades productivas para validación de modelo

Tabla V. **Definiendo tamaño de la muestra**

	<b>Unidades productivas para entrenamiento del modelo</b>	<b>Unidades productivas para validación del modelo</b>
Población	Número	Número

Fuente: elaboración propia, empleando Microsoft Excel 365.

#### **10.1.5. Error de muestreo**

Es importante determinar el error de la muestra para validar si la muestra es representativa o no, esto aplicará tanto a los datos iniciales de la investigación, como a lo largo del desarrollo de esta, como por ejemplo al determinar el tamaño de muestra para entrenamiento y validación del modelo, será importante validar que el error de muestreo dé la fiabilidad de que la muestra será representativa para la investigación.

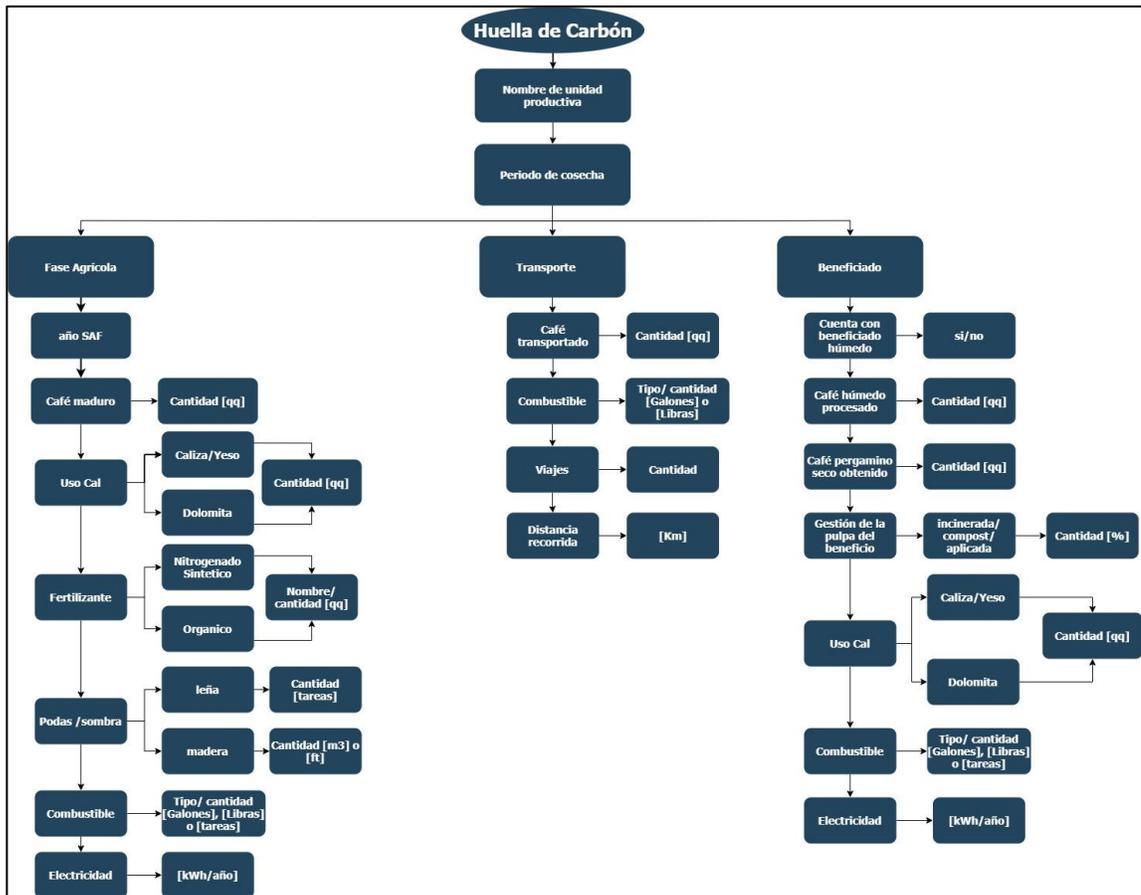
#### **10.2. Análisis de la información**

Es momento de proceder con el análisis de la información, en este punto ya se conocen los datos con los cuales se llevará a cabo el entrenamiento y validación del modelo de datos. Para llevar a cabo la etapa de análisis se hará uso de técnicas de estadística descriptiva inferencial

### **10.2.1. Variables del proceso**

Es importante determinar todas las variables que estarán involucradas dentro del modelo para determinar la cantidad de huella de carbono que se emite en las distintas unidades productivas del sector SAF café, por lo cual por medio de un diagrama de bloques en cascada se plasmó las variables que se consideran importantes dentro del sistema. A continuación, se muestra dicho diagrama indicando las etapas y las variables correspondientes a cada etapa. Como se podrá observar, existe una gran cantidad de variables a considerar dentro del proceso, y cada una de estas contribuye a las emisiones de GEI en distintas proporciones, sin embargo, ninguna de ellas es despreciable ya que, aunque sea en mínima cantidad emite una cantidad sustancial de CO<sub>2</sub> y todas deben tomarse en cuenta en todo momento.

Figura 13. Variables del proceso



Fuente: elaboración propia, realizado con Visio 2019.

### 10.2.2. Análisis de tendencia central

El análisis de tendencia central se llevará a cabo por medio la estadística descriptiva utilizando técnicas como la media, moda y mediana para observar que datos tienen mayor influencia sobre la emisión de CO<sub>2</sub> dentro de la producción de café, estas posteriormente serán contrastadas contra otros métodos para identificar cual tiene el acercamiento más acertado y útil para esta

investigación, estos métodos se llevaran a cabo por medio del software estadístico RStudio.

A la vez, se podrá utilizar estas técnicas para mostrar cuantas unidades productivas se encuentran en determinados rangos de CO<sub>2</sub> emitido.

### **10.2.3. Análisis de dispersión**

Por medio de la estadística descriptiva las medidas de dispersión brindaran un soporte muy importante dando visión respecto a la variación que existe entre las variables de estudio (huella de carbono, agua, energía eléctrica, entre otras) mostrando de una forma más entendible la dispersión o distancia que existe dentro del conjunto de datos de cada tipo respecto a la emisión de CO<sub>2</sub>. Con este dato se afianzan los resultados obtenidos en el análisis de tendencia central de los datos.

### **10.2.4. Comparación entre variables**

Es el momento de utilizar las técnicas de estadística inferencial, esta vez por medio de las herramientas T-Student y ANOVA, se llevará a cabo el análisis comparativo de datos para las variables o grupos de estudio seleccionados, por ejemplo, al momento de comparar la media de CO<sub>2</sub> emitido por unidades productivas o la emisión de CO<sub>2</sub> por el uso de fertilizantes entre dos unidades productivas.

### **10.2.5. Comportamiento entre variables**

Al igual que en el punto 10.2.4 se utiliza técnicas de estadística inferencial, pero en esta ocasión se recurrirá al estudio de regresión entre las variables, esto

para determinar la relación o cambios que puedan existir entre dos o más variables del sistema de producción de SAF café durante el periodo de tiempo propuesto, por medio del coeficiente de correlación  $r$ , para ello se utilizará métodos como Chi-cuadrado o Pearson. Este método se aplicará por medio del software estadístico RStudio.

### **10.2.6. Estudio del comportamiento de variables**

El método de regresión del punto 10.2.5 será fundamental dentro de este trabajo de investigación, ya que una vez se haya establecido el nivel de asociación que existe entre las distintas variables y se haya determinado que método de correlación es el más apto, será el momento de proceder con el estudio del comportamiento de las variables tomando como base inicial del proyecto la ecuación básica del modelo de regresión lineal simple que se muestra a continuación:

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

Este método se aplicará por medio del software estadístico como RStudio.

### **10.3. Presentación de los resultados**

La presentación de resultados es la etapa donde se mostrarán los resultados obtenidos de los procesos anteriores (preparación y análisis de la información), esto se llevará a cabo por medio de:

- Tablas de datos
- Gráficos de datos

### 10.3.1. Tablas de datos

- Los resultados correspondientes a la cantidad de CO<sub>2</sub> emitido por unidad productiva se mostrarán en una tabla de datos similar a la mostrada a continuación. En ella se detallará la variable independiente que actúa sobre la emisión de CO<sub>2</sub>, mostrando la cantidad y porcentaje emitido respectivamente.

Tabla VI. **Ejemplo cantidad de CO<sub>2</sub> emitido por unidad productiva**

<b>Nombre de unidad productiva</b>		
<b>Emisor</b>	CO <sub>2e</sub>	%CO <sub>2e</sub>
<b>Electricidad</b>	Valor	%valor
<b>Combustible</b>	Valor	%valor
<b>Cal</b>	Valor	%valor
<b>Pulpa</b>	Valor	%valor
<b>...</b>	Valor	%valor
<b>Total</b>	Valor	100 %

Fuente: elaboración propia, empleando Microsoft Excel 365.

- A la vez, se muestra un ejemplo de la presentación de los datos para comparar las emisiones de CO<sub>2</sub> por cada una de las variables independientes, y como afectan a cada una de las unidades productivas, dentro de la tabla de datos se podrá observar el porcentaje de cada variable independiente respecto del total de emisiones.

Tabla VII. **Ejemplo comparación de CO2 emitido por unidad productiva**

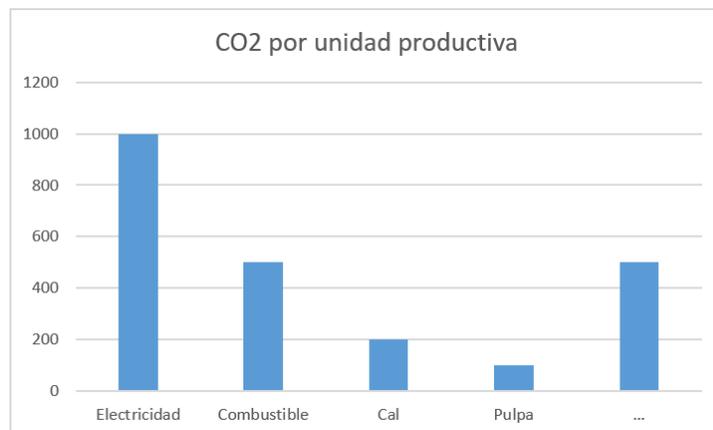
Unidad productiva	Variable de emisión de CO <sub>2</sub>					Total
	Electricidad	Combustible	Cal	Pulpa	...	
<b>1</b>	valor	valor	valor	valor	valor	valor
<b>2</b>	valor	valor	valor	valor	valor	valor
<b>3</b>	valor	valor	valor	valor	valor	valor
<b>4</b>	valor	valor	valor	valor	valor	valor
<b>Total</b>	valor	valor	valor	valor	valor	valor
<b>%Total</b>	%valor	%valor	%valor	%valor	%valor	100 %

Fuente: elaboración propia, empleando Microsoft Excel 365.

### 10.3.2. Gráficos de datos

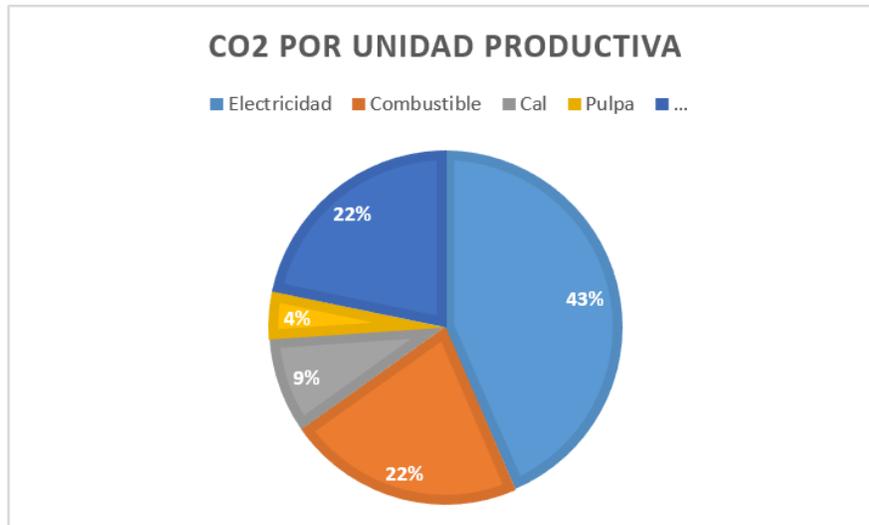
El uso de técnicas de visualización de datos como diagramas de dispersión, histogramas, barras de datos, entre otros, se utilizarán para mostrar de una forma gráfica el análisis de los datos obtenidos, como por ejemplo la relación que existe entre la cantidad de CO<sub>2</sub> producido en función de la altitud de una región, o si se habla de una distribución normal de datos.

Figura 14. **Ejemplo 1 de representación gráfica de datos**



Fuente: elaboración propia, empleando Microsoft Excel 365.

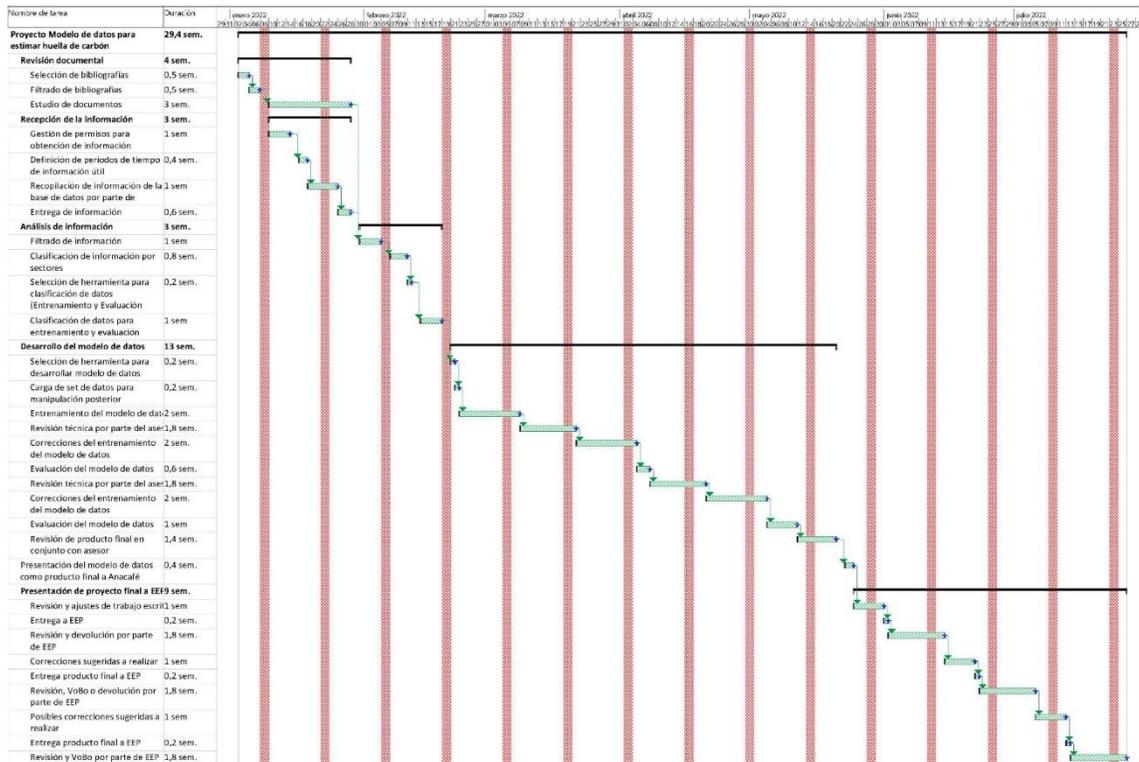
Figura 15. **Ejemplo 2 de representación gráfica de datos**



Fuente: elaboración propia, empleando Microsoft Excel 365.

# 11. CRONOGRAMA

Figura 16. Cronograma de actividades propuestas



Fuente: elaboración propia, empleando Microsoft Project 2019.



## 12. FACTIBILIDAD DEL ESTUDIO

El presente trabajo de investigación cuenta con el apoyo de la Asociación Nacional del Café –Anacafé-, quien en pro del conocimiento y de la cultura de producción más limpia, contribuirá con datos asociados a diferentes áreas de cultivo distribuidas en el territorio nacional.

### 12.1. Factibilidad Operativa

- Recurso humano

Sera necesario:

- 1 persona encargada de recabar la información necesaria para la alimentación de la base de datos con la cual se desarrollará el trabajo de investigación.
- 1 persona del departamento de ambiente que contribuya con conocimientos técnicos referentes al ámbito del café, GEI, entre otros.
- Personal con conocimientos en el área química o de laboratorio para para brindar soporte en temas de suelo, aguas u otros factores.
- Personal que se encuentre directamente involucrado dentro de los procesos productivos de las plantaciones de SAF café.
- 1 persona que llevará a cabo el modelo de datos.

- **Periodo de datos**
  - El SAF café en Guatemala consta de un periodo de tiempo no muy extenso para llevar a cabo la primera fase del proceso de producción, siendo esta la fase de la cosecha que inicia a finales de noviembre y finaliza a inicios del mes de abril.
  - Se procederá a seleccionar un periodo de datos de al menos 2 años productivos del SAF café para contar con la data necesaria para el respectivo análisis.
  - Se analizará si el impacto mundial causado por la pandemia SARS COVID-19 tiene influencia representativa que altere los datos de estudio o si es despreciable en la toma de decisiones.
  
- **Acceso a la información**
  - El acceso a la información correspondiente al flujo de la producción del cultivo del SAF café es primordial para llevar a cabo este trabajo de investigación.
  
- **Permisos**
  - Es de vital importancia contar con los permisos necesarios para poder acceder a las bases de datos de información relevante al tema de la huella de carbono de Anacafé.

## **12.2. Factibilidad técnica**

- **Aspectos técnicos de Hardware y Software**
  - Prestaciones mínimas de equipo de cómputo a utilizar
    - Procesador Intel Core i5 3era. Generación
    - 4 GB de RAM
    - 512 GB de almacenamiento

- Software
  - OS Windows 10
  - R
  - R Studio
  - Python
  
- Conocimientos del usuario de pruebas técnicas
  - Conocimiento en el proceso de producción del SAF café y sus respectivas etapas.
  - Conocimiento sobre el uso específico del modelo de datos.
  - Conocimiento básico acerca de modelación de datos.
  - Lenguaje técnico intermedio.
  
- Conocimientos del usuario final
  - Conocimiento en el proceso de producción del SAF café y sus respectivas etapas.
  - Conocimiento sobre el uso básico del modelo de datos.
  - Lenguaje técnico no indispensable.
  
- Periodo de tiempo para la creación del modelo
  - El modelo planteado en la actual investigación se proyecta con fecha enero 2022 como punto de partida, y luego de un trabajo de aproximadamente 7 meses, se espera la entrega del producto final con fecha julio 2022.

### 12.3. Factibilidad económica

La factibilidad económica del proyecto es parte fundamental para poder obtener los resultados esperados de la investigación, es por esto que, en la siguiente tabla se indica los posibles costos asociados al desarrollo del modelo.

Tabla VIII. **Datos de factibilidad económica**

<b>Actividad</b>	<b>Costo (Q)</b>	<b>Detalle</b>
<b>Recurso humano</b>	0	El recurso humano que fungirá como apoyo al trabajo de investigación, llevará a cabo sus actividades Ad-Honorem.
<b>Uso de software</b>	0	El software por utilizar será de uso libre, por lo cual no existirá ningún costo asociado a las actividades involucradas del mismo.
<b>Uso de hardware</b>	0	El equipo de cómputo utilizado para el desarrollo y modelado de datos será proporcionado por el investigador, por lo cual no incurre en ningún costo adicional.
<b>Elaboración del modelo de datos</b>	0	La elaboración del modelo de datos corresponde al objetivo principal de esta investigación, la cual se desarrolla en aras de la educación y obtención de conocimiento, por lo cual, no tendrá ningún costo asociado y se realizará de forma gratuita.
<b>Otros costos asociados</b>	0	No existirán desplazamientos o actividades extra que ameriten un costo a la realización de esta investigación.

Fuente: elaboración propia, empleando Microsoft Excel 365.

Como se pudo observar en la tabla previamente detallada, el presente trabajo de investigación y desarrollo del modelo de datos no amerita llevar a cabo ningún gasto por parte de los involucrados.

Al cumplir con los elementos técnicos, operativos y económicos necesarios, se puede concluir que la elaboración del modelo de datos propuesto en este trabajo de investigación y las actividades relacionadas con el mismo son factibles a desarrollar.

## 13. REFERENCIAS

1. Agencia de protección ambiental de Estados Unidos. (23 de junio de 2021). Emisiones de dióxido de carbono en EE-UU en 2017. [Mensaje de blog]. Recuperado de <https://espanol.epa.gov/la-energia-y-el-medioambiente/emisiones-de-dioxido-de-carbono>.
2. Aguirre, A. y Garay, K. (2021). *Determinación de la huella de carbono en el cultivo y procesamiento del café y estrategias para su reducción* (Tesis de maestría). Universidad Internacional SEK, Quito.
3. Albornoz, A. (2017). *Huella de Carbono del Café (Coffea arabica) en Empresa Asociativa Campesina Aruco en Copán* (Tesis de licenciatura). Escuela Agrícola Panamericana, Honduras.
4. Anacafé. (2019). *Guía ambiental para el sector café de Guatemala*. Guatemala: Ministerio de Ambiente y Recursos Naturales.
5. Anacafé. (2020). *Memoria de labores 2019 - 2020*. Guatemala: Ministerio de Ambiente y Recursos Naturales.
6. Andrade, Hernán J.; Marín, Lina M.; Pachón, Diana P. (agosto, 2014). Fijación de carbono y porcentaje de sombra en sistemas de producción de café (*Coffea arabica* L.) en el Líbano, Tolima, Colombia. *Revista Bioagro*, 26(2), 127-132.

7. Ariza, W., Arias, J., Riaño, N., Riaño, A., Posada, H., Valenzuela, J., Vega, M., Murgueitio, Y. y Castro, J. (junio, 2018). Determinación de la huella de carbono en el sistema de producción de café pergamino seco, de cuatro municipios del sur del departamento del Huila (Colombia). *Revista De Investigación Agraria Y Ambiental*, 9(2), 109 - 120.
8. Asociación Nacional del Café -Anacafé-. (2016). *Política de Ambiente y Cambio Climático para el sector café de Guatemala*. Guatemala: Anacafé.
9. De la Puente, C. (2018). *Estadística Descriptiva e Inferencial*. Madrid, España: Ediciones IDT.
10. Gamboa, M. (2017). *Estadística aplicada a la investigación científica*. Las Tunas, Cuba: Editorial Redipe-Edacun.
11. Gavilán, E. y Reinoso, M. (2016). *Estimación cuantitativa de la huella del carbono en el cultivo de la caña de azúcar en Villa Clara* (Tesis de licenciatura). Villa Clara, Cuba.
12. GreenFacts. (18 de mayo de 2021). Captura y Almacenamiento de CO2. [Mensaje de blog]. Recuperado de <https://www.greenfacts.org/es/captura-almacenamiento-co2/index.htm#4>.
13. James, G., Witten, D., Hastie T. y Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*, New York, USA: Springer.

14. Kaplan, D. (agosto, 2018). Teaching Stats for Data Science. *The American Statistician*, 72(1), 89-96, DOI: 10.1080/00031305.2017.1398107.
15. Mora, A. y Mendoza, S. (2017). *Determinación de impactos ambientales en la finca Cascajal en Pacho Cundinamarca, mediante el balance de carbono y nutrientes para un cultivo de café* (Tesis de licenciatura). Universidad de la Salle, Colombia. Recuperado de [https://ciencia.lasalle.edu.co/ing\\_ambiental\\_sanitaria/356](https://ciencia.lasalle.edu.co/ing_ambiental_sanitaria/356).
16. Olmos, X. (2021). *Guía básica para la recolección y validación de los datos necesarios para calcular la huella ambiental del café verde según la norma europea*. Chile: División de Comercio Internacional e Integración, CEPAL.
17. Quirós, A. (2017). *Inventarios de GEI y Huella de carbono para Beneficios*. Costa Rica: Cooperación Alemana para el Desarrollo (GIZ).
18. Sanchez, M., Cabrera, M., Rojas, F., Ortiz, J., Gordillo, S. y Perdomo, D. (septiembre, 2017). Software para el cálculo de la huella ambiental en la producción de cacao. *Memorias De Congresos UTP*, 173-179. Recuperado a partir de <https://revistas.utp.ac.pa/index.php/memoutp/article/view/1488>.
19. Segura, M. y Andrade, H. (2015). *Huella de carbono en cadenas productivas de café (Coffe arabica L.) con diferentes estándares de certificación en Costa Rica* (Tesis de licenciatura). Universidad de Caldas, Costa Rica.

20. Wickham, H. y Golemund, G. (2017). *R for data science: Import, tidy, transform, visualize, and model data*. Canada: O'Reilly Media, Inc.
21. Zaldaña, F. (2019). *Determinación de emisiones de gases de efecto invernadero y alternativas para la gestión eficiente de captura de carbono a través de la herramienta cool farm tool en los sistemas de producción de Coffea sp. en Rodríguez de Mendoza - Amazonas 2018* (Tesis de licenciatura). Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas, Chachapoyas, Peru.