**Documento Final del Proyecto: Análisis de Variables que Influyen en la Tasa de Extracción de Aceite (TEA) – Grupo Palmas**

**1. Introducción**

Este proyecto se realizó con el objetivo de identificar y analizar las variables que influyen en la Tasa de Extracción de Aceite (TEA) en las 4 plantaciones principales de Grupo Palmas:

* Palmawasi
* Shanusi
* Tulumayo
* Nuevo Horizonte.

A través de un enfoque basado en análisis descriptivo y la construcción de un modelo de pesos de las variables, se desarrollaron dos fases que permitieron tanto la identificación de patrones como la cuantificación de la influencia de dichas variables.

El estudio se llevó a cabo utilizando Python versión 3.9 y el repositorio oficial del proyecto con el código completa se encuentra en la siguiente ruta de GitHub: <https://github.com/cesarqb/PROYECTO_01_analitica_descriptiva_TEA>

**2. Objetivos**

* **Fase 1:** Identificar patrones, tendencias y relaciones significativas entre las variables que influyen en la variabilidad de la TEA.
* **Fase 2:** Determinar los pesos de influencia de las variables identificadas en la Fase 1 sobre la TEA para cada una de las plantaciones, generando simulaciones con Power BI.

**3. Metodología**

El análisis fue desarrollado en dos fases principales:

**3.1 Fase 1: Análisis Descriptivo**

Esta fase se centró en la exploración y análisis descriptivo de las variables que podrían influir en la TEA, tales como la cantidad de fruta fresca, fruta sobremadura, frutas dañadas, clima y condiciones del suelo. El objetivo fue identificar patrones y relaciones significativas entre las variables y la TEA.

El análisis incluyó:

* Visualización de las tendencias de la TEA a lo largo del tiempo.
* Detección de posibles correlaciones entre las variables.
* Identificación de outliers y anormalidades en los datos.

Los resultados finales se encuentran en la ruta:

<https://github.com/cesarqb/PROYECTO_01_analitica_descriptiva_TEA>

**3.2 Fase 2: Modelado y Simulación de Pesos**

En esta etapa, se implementó un modelo que permitió calcular los pesos de las variables más influyentes sobre la TEA. Este modelo se alimentó de los insights obtenidos en la primera fase y fue desarrollado para cada una de las 4 plantaciones.

El objetivo era proporcionar a los stakeholders una simulación de cómo los cambios en estas variables afectan directamente la TEA.

**Inicio de la Fase 2:** 01 de agosto de 2024.

**4. Herramientas Utilizadas**

* **Lenguaje de Programación:** Python 3.9
* **Plataforma de Control de Versiones:** GitHub
* **Librerías:** Se utilizaron librerías como Pandas, NumPy, Matplotlib y Scikit-learn, entre otras, para el análisis de datos y modelado.
* **Visualización y Simulaciones:** Power BI fue empleado para visualizar los resultados y realizar simulaciones basadas en los pesos de las variables.

**5. Estructura del Repositorio GidHub del proyecto:**

El repositorio está estructurado en las siguientes carpetas clave, facilitando la organización y flujo del proyecto:

├── data/

│ ├── raw/ -> Data cruda, datos descargados sin modificar

│ ├── external/ -> Data de fuentes de 3eras partes

│ ├── interim/ -> Datos intermedios que han sido transformados

│ └── processed/ -> Datos procesados, que serán usados para el modelo

├── src/

│ ├── clean\_scripts/ -> Scripts para cargar, limpiar, transformar y procesar datos

│ ├── features\_scripts/ -> Scripts para generar características a partir de los datos

│ ├── model\_scripts/ -> Modelos de aprendizaje automático y scripts

├── informes/

│ ├── informe\_analitica\_descriptiva/ -> Archivo PPT mostrando los hallazgos mas importantes a nivel de variables detectadas

│ ├── informe\_analisis\_pesos/ -> Archivo PPT mostrando los pesos y la relación de importancia de variables sobre cada plantación

├── requirements.txt -> Requerimientos de versiones de librerias necesarias para el proyectos

* **Data:**
  + **external/:** Almacena los datos originales descargados de diferentes fuentes.
  + **processed/:** Almacena los datos que han sido limpiados y transformados para el análisis usando el programa python.
* **Src:**
  + **clean\_scripts/**: [Scripts de limpieza](https://github.com/cesarqb/PROYECTO_01_analitica_descriptiva_TEA/tree/main/src/clean_scripts), contiene scripts de Python utilizados para la limpieza y transformación de los datos originales.
  + **features\_scripts/**: Scripts para la creación de características y consolidación de análisis por sede.
  + **model\_scripts/:** Scripts utilizados en la segunda etapa del proyecto para determinar los pesos de las variables por plantación.
* **Informes:**
  + **informe\_analitica\_descriptiva/:** Informe en formato PPT con los hallazgos más importantes de la Fase 01 del Proyecto.
  + **informe\_analisis\_pesos/:** Informe en formato PPT con los hallazgos más importantes de la Fase 02 del Proyecto.

**Informe cierre proyecto Analitico \_status/** Informe en formato PPT con el resumen de objetivos, entregables, hallazgos y pendientes de la Fase 01 y Fase 02 de los proyectos analíticos realizados

**6. Resultados**

* En la **Fase 1**, se detectaron patrones claros que vinculaban la cantidad de fruta fresca y la fruta sobremadura como las variables más determinantes para la TEA en cada plantación.
* En la **Fase 2**, se lograron determinar los pesos de influencia de cada variable para cada una de las 4 plantaciones, lo que permitió simular escenarios de producción y cómo la TEA varía dependiendo de los cambios en las variables controlables, como las condiciones de la fruta.

**7. Entregables**

1. **Informes Finales:** Se presentaron informes detallados a los stakeholders, explicando los hallazgos del análisis y las simulaciones de la Fase 2.
2. **Dashboard en Power BI:** Un dashboard que permite interactuar con las simulaciones de los pesos de las variables, mostrando de manera dinámica cómo la TEA se ve afectada por diferentes factores.
3. **Repositorio GitHub:** Todo el código, los datos procesados y las simulaciones están disponibles en el [repositorio oficial del proyecto](https://github.com/cesarqb/PROYECTO_01_analitica_descriptiva_TEA).

**8. Stakeholders**

El equipo de stakeholders estuvo conformado por:

* Zeyda Alfaro (Jefe de investigación)
* César Díaz (Jefe de transformación digital y BI)
* Luis Fernando Díaz (Project Leader)
* Juan Enrique Córdova (Lider TMO)
* Fernando Enrique Núñez (Analista de estrategia y control)
* Rocky Raúl Lizama (Super intendente de ATC)
* Juan Carlos Chávez (Gerente agrícola)

El equipo de analistas que participó en el proyecto incluye:

* Roly Cruz (Consultor BI)
* César Quezada (Consultor BI)

**10. Conclusión**

Este proyecto ha permitido a Grupo Palmas obtener una visión más profunda y basada en datos de los factores que influyen en la TEA. La implementación del modelo de pesos y las simulaciones no se han concluido pero el seguimiento de los mismos proporcionará a la empresa una herramienta valiosa para la toma de decisiones estratégicas relacionadas con la optimización de su producción.

**11. Recomendaciones**

* Continuar monitoreando las variables clave y actualizar el modelo con nuevos datos de manera periódica para refinar los pesos y mejorar la precisión de las simulaciones.
* Implementar las simulaciones en los procesos de planificación para prever el impacto de las fluctuaciones en las variables más influyentes sobre la TEA.