



ESCUELA AGRÍCOLA PANAMERICANA- ZAMORANO-
TENDENCIAS E INNOVACIÓN EN TECNOLOGÍA AGRÍCOLA
PH.D. SERVIO PALACIOS / ING. GONZALO MARADIAGA

Proyecto Final CAPSTONE

NUTRI- ANALYTICS

EQUIPO: GOLD HEALTH



SEBASTIAN SALVATIERRA 24278
FRANCIS RÍOS 24174
ROBERTO MINUCHE 24051
IZZA MOREIRA 24151
VICTOR POSADAS 24215
JORGE MÉNDEZ 24283
SARA MACH 24167
KIMBERLY ARAGÓN 24210
DANIA PEINADO 24137
MELISSA GONZALEZ 24253
VALERIA ARAUJO 24277

LÍDER: MELISSA GONZALEZ 24253

ENFOQUE DEL PROYECTO: CREACIÓN DE GRÁFICOS PARA ANALIZAR ÍNDICE DE MASA CORPORAL DE LA POBLACIÓN ESTUDIANTEL DE ZAMORANO DE LAS CLASES 2023, 2024 Y 2025.

**SAN ANTONIO DE ORIENTE, FRANCISCO MORAZÁN, HONDURAS
NOVIEMBRE DE 2022**

I. Problema

En la actualidad se utilizan múltiples métodos de recolección de información en el ámbito nutricional para evaluar el estado de una población con el fin de tomar acciones preventivas y correctivas sobre su situación. Sin embargo, estos métodos exponen carencias tecnológicas que dificultan el análisis de los datos. Aunado a ello, este es un tema tratado a nivel internacional, puesto que, existen grandes afecciones relacionadas con la alimentación y actividad física de la población estudiantil de Zamorano, destacándose entre ellas la desnutrición, malnutrición, sobrepeso y obesidad, es por ello, que el objetivo principal de este proyecto es la creación de gráficos mediante el uso de Jupiter-Python y Matplotlib sobre el índice de masa corporal y correlacionar los datos de estatura con el peso esto con la finalidad de determinar el estado nutricional en que se encuentran los estudiantes.

II. Objetivos

El proyecto incluye los siguientes objetivos:

- Crear un repositorio público de código fuente para el proyecto: Nutri-Analytics.
- Automatizar la ingestión de una fuente de datos CSV de datos de peso, edad y estatura de la población estudiantil de Zamorano mediante el software Python.
- Digitalizar de manera automática la subida a memoria de los datos en un formato consistente para facilitar el análisis de los datos recolectados.
- Analizar gráficas que faciliten la interpretación del índice de masa corporal de la población estudiantil de Zamorano.
- Correlacionar el Índice de Masa Corporal de la población con respecto a su estatura y la estatura con el peso.

III. Solución

Para este proyecto, se realizó el uso del lenguaje de programación Python, el cual fue de ayuda para automatizar la inspección de datos sobre los diferentes IMC'S de las diferentes poblaciones de Zamorano, que fueron recolectados mediante el software de administración de encuestas, Google Forms, además se empleó el editor de código fuente de Jupiter-Python, mediante el uso de la librería Matplotlib. Lo anteriormente mencionado, es un todo que nos permitieron desarrollar un programa que aprovechó los datos obtenidos de la población estudiantil y los interpretó para darles así, el enfoque estadístico.

IV. Herramientas y Métodos

i. Herramientas:

- Jupyter
- Github
- Encuesta de google forms

- Lenguaje de programación Python
- Microsoft Excel 365

ii. Librerías:

- Pandas
- Matplotlib
- Numpy
- Statistics
- Sklearn

iii. Métodos:

Para la resolución del problema, primeramente, se realizó una encuesta de Google Forms a los estudiantes de primero, segundo y tercer año, en la cual, se les preguntó los siguientes puntos: edad, peso (lb), estatura (m) y año de carrera.

Los datos obtenidos de la encuesta se tabularon en Microsoft Excel para ordenar los datos y poder analizarlos posteriormente en jupyter notebooks con ayuda de librerías y Python. Se utilizó en formato CSV.

Python

Se importaron librerías, las cuales son:

- **Sklearn:** es una librería para modelo de regresión lineal.
- **Pandas:** es una herramienta de manipulación y análisis de datos de código abierto rápida, potente y flexible de usar, construida sobre el lenguaje de programación Python.
- **Matplotlib:** es una biblioteca completa para crear visualizaciones estáticas, animadas e interactivas en Python. Se utilizó para realizar las gráficas de los datos analizados.
- **Numpy:** en la biblioteca de Python que se utilizó para trabajar las operaciones matemáticas.
- **Statistics:** Esta librería nos permite realizar cálculos estadísticos para nuestro análisis.
- **Estadístico:** se utilizó para el análisis estadístico y con el uso del programa Survey Monkey se estimó el tamaño de la muestra <https://es.surveymonkey.com/mp/sample-size-calculator/>

Gráficas:

En este proyecto se hizo uso de diferentes herramientas de desarrollo como lo son Python y Visual Studio Code, con la finalidad de analizar los datos. Aunado a esto, se implementó GitHub como

repositorio de código fuente del proyecto el cual es público para que, de esta forma, la audiencia tenga acceso a la información.

Las gráficas que se realizaron para analizar los datos son las siguientes:

- Gráfico de barras del peso promedio por año
- Gráfico de líneas del efecto del género sobre el peso de cada clase
- Gráfico de regresión y correlación lineales simple de peso vs estatura.
- Gráfico de regresión y correlación lineales simple del peso vs IMC
- Histograma de primer, segundo y tercer año.
- Gráfico de Barras apilada Población, estado IMC

V. Resultados

i. Fuente de datos publicados en Github

La Fuente de Datos recopilada y utilizada está presente en el siguiente link:
<https://github.com/GoldHealth/Nutri-Analytics/blob/main/data/vimc.csv>

ii. Estructuras de datos utilizadas

Es una Base de Datos, en formato .csv, el cual, es un formato simple y fácil de manipular.

iii. Código fuente publicado en Github

El código fuente de cada módulo de gráficos está presente en el siguiente enlace de Github:

<https://github.com/GoldHealth/Nutri-Analytics/tree/main/plot%20drafts>

iv. Código fuente documentado

A.- Histograma 1er, 2do, 3er Año: <https://github.com/GoldHealth/Nutri-Analytics/tree/main/plot%20drafts>

B.- Grafico de Barras, Peso Promedio por Clase: [https://github.com/GoldHealth/Nutri-Analytics/blob/main/plot%20drafts/Grafico%20de%20Barras-Peso%20Promedio%20por%20clase%20\(1\).ipynb](https://github.com/GoldHealth/Nutri-Analytics/blob/main/plot%20drafts/Grafico%20de%20Barras-Peso%20Promedio%20por%20clase%20(1).ipynb)

C.- Grafico de Líneas Genero-Peso: [https://github.com/GoldHealth/Nutri-Analytics/blob/main/plot%20drafts/Grafico%20de%20lineas%20efecto%20del%20genero%20sobre%20el%20peso%20en%20clase%20\(1\).ipynb](https://github.com/GoldHealth/Nutri-Analytics/blob/main/plot%20drafts/Grafico%20de%20lineas%20efecto%20del%20genero%20sobre%20el%20peso%20en%20clase%20(1).ipynb)

D.- Gráfico de Barras apilada Población-estado IMC: <https://github.com/GoldHealth/Nutri-Analytics/blob/main/plot%20drafts/Grafico%20de%20Barras%20apilado>

E.- Regresión, Correlación lineal de Estatura-Peso: <https://github.com/GoldHealth/Nutri->

[Analytics/blob/main/plot%20drafts/Grafico%20de%20regresion%20Lineal%20y%20Correlacion%20lineal%20simple_EstaturavsPeso.ipynb](https://github.com/GoldHealth/Nutri-Analytics/blob/main/plot%20drafts/Grafico%20de%20regresion%20Lineal%20y%20Correlacion%20lineal%20simple_EstaturavsPeso.ipynb)

F.- Regresión, Correlación lineal de Peso-IMC: [https://github.com/GoldHealth/Nutri-Analytics/blob/main/plot%20drafts/Grafico%20de%20regresion%20Lineal%20y%20Correlacion%20lineal%20simple_PesovsIMC%20\(1\).ipynb](https://github.com/GoldHealth/Nutri-Analytics/blob/main/plot%20drafts/Grafico%20de%20regresion%20Lineal%20y%20Correlacion%20lineal%20simple_PesovsIMC%20(1).ipynb)

v. Descripción de las funciones y módulos utilizados

A.- Histograma 1er, 2do, 3er Año: Se utilizo el histograma para observar el comportamiento de los datos. Este módulo pretende tomar todos los datos y ver si su comportamiento es normal. Una herramienta fundamental fue la librería statistics para poder hallar la desviación estándar y la media, asimismo, permite tener el error estándar, el cual, se usa para plotear el histograma.

B.- Gráfico de Barras, Peso Promedio por Clase: importamos numpy para hallar las medias del peso para cada año, puesto que, matplotlib suele graficar los valores más altos. Por ello, se extrajeron dataframes seleccionados por año cursado para la media, para posteriormente utilizar matplotlib para graficar en barras horizontales.

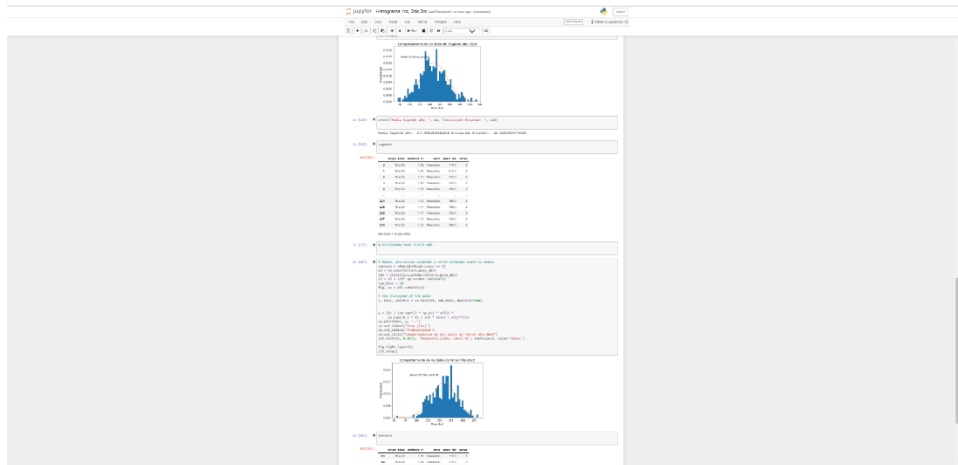
C.- Gráfico de Líneas Genero-Peso: Este nos permite observar cómo se comporta el peso de los estudiantes acorde al avance de la clase que pertenecen y a su género. Básicamente, utilizamos las medias de cada año, para cada sexo, y una media general por año.

D.- Gráfico de Barras apilada Población-estado IMC: Nos permite, comparar porcentualmente, la clasificación por cada clase, respecto al IMC, según las categorías, Bajo Peso, Peso Normal, Sobrepeso y Obesidad, esto con los estándares adaptados de <https://www.revista-portalesmedicos.com/revista-medica/wp-content/uploads/SOBREPESO-Y-LA-OBESIDAD.pdf>

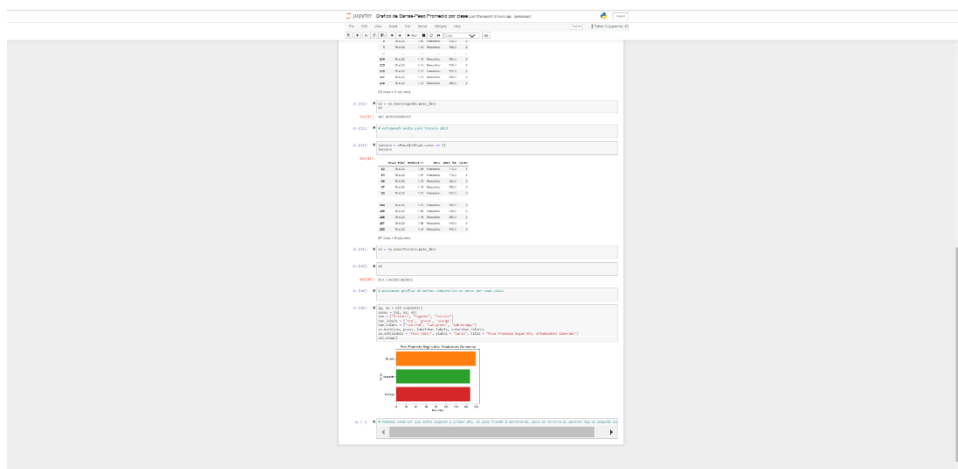
E.- Regresión, Correlación lineal de Estatura-Peso y Peso-IMC: Este módulo permite realizar un modelo lineal predictivo para establecer el efecto de la estatura por sobre el peso y el peso sobre el IMC. Aquí se utilizó la librería sklearn con su función linear_model, este permite utilizar la fuente de datos para directamente plotearlo, se utilizó y modifíco una fuente externa para realizar este módulo.

vi. Pruebas de código mostrando el funcionamiento correcto del código fuente y demostración de la salida del código

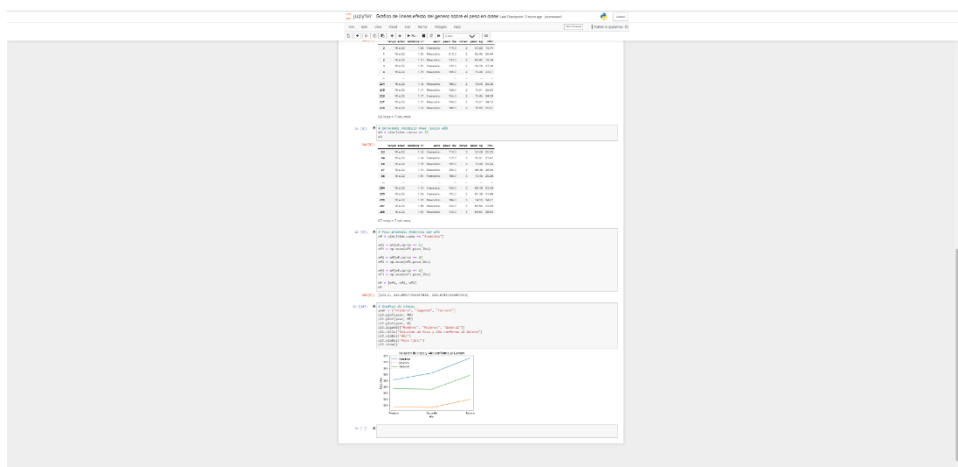
A.- Histograma 1er, 2do, 3er Año:



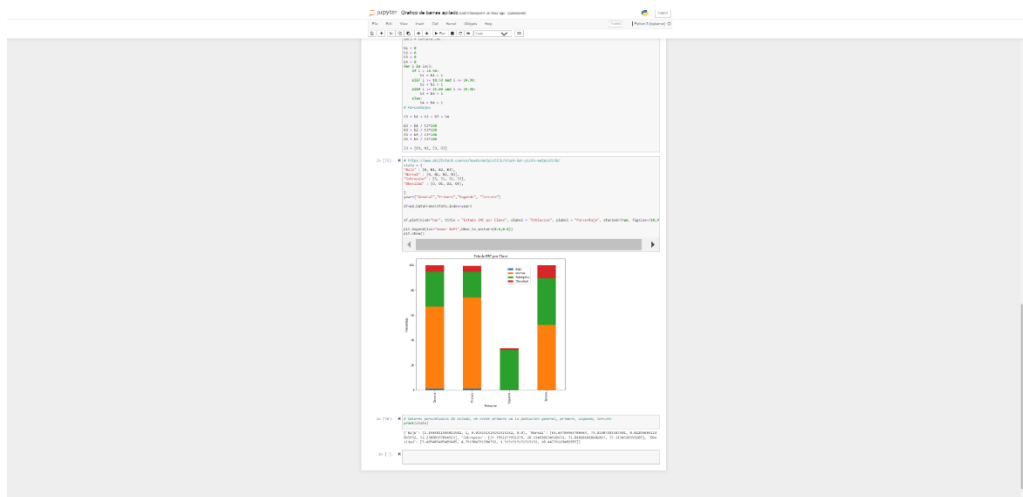
B.- Grafico de Barras, Peso Promedio por Clase:



C.- Grafico de Líneas Genero-Peso:

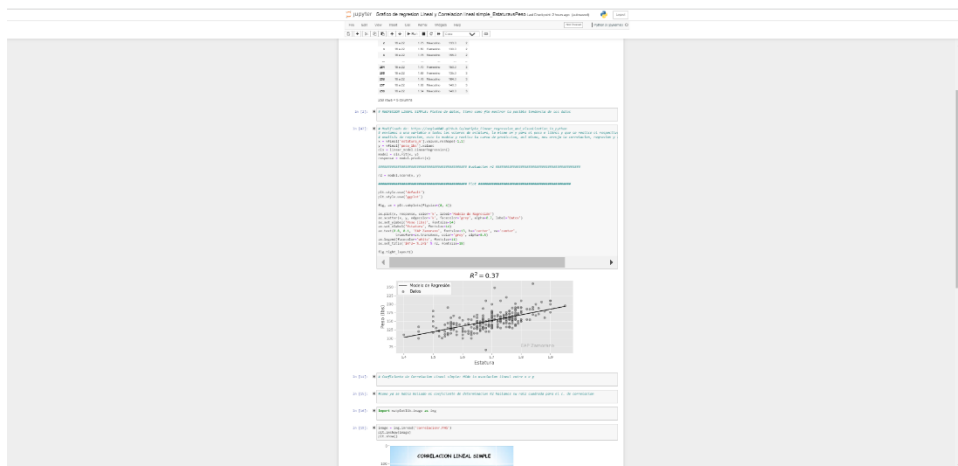


D.- Gráfico de Barras apilada Población-estado IMC:

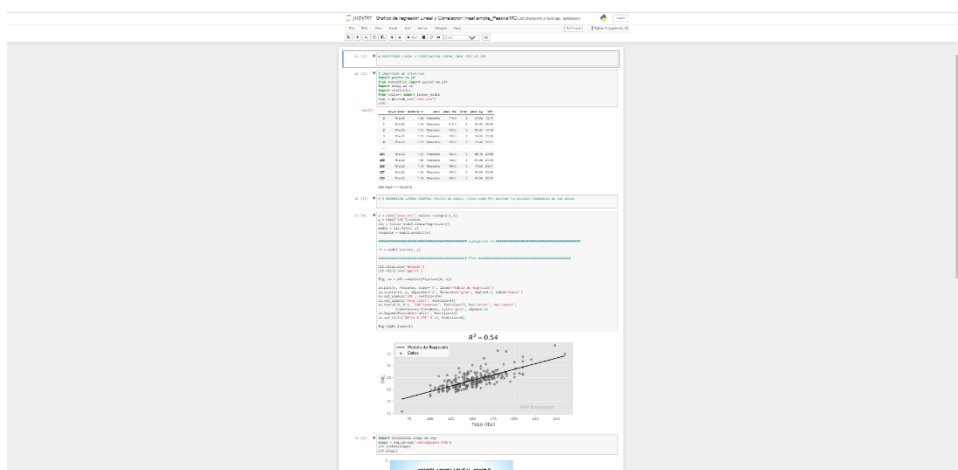


E.- Regresión, Correlación lineal de Estatura-Peso y Peso-IMC:

- Estatura-peso



-Peso-IMC



Note: Si desea ver detalladamente los códigos, visitar el siguiente enlace en Github: <https://github.com/GoldHealth/Nutri-Analytics/tree/main/plot%20drafts>

Presentación de Resultados

Histograma 1er, 2do, 3er Año:

El histograma presentado a continuación denota que los datos intentan encajar en una trayectoria similar a la “Campana de Gauss”, por lo cual se puede determinar mediante esta prueba grafica que en efecto existe una distribución normal en los pesos de la población estudiantil de la clase 2025. Además, se obtuvo una desviación estándar de 24.33% indicando con ello que la mayor parte de los datos de la muestra tienden a estar agrupados cerca de su media equivalente a 148.36 libras.

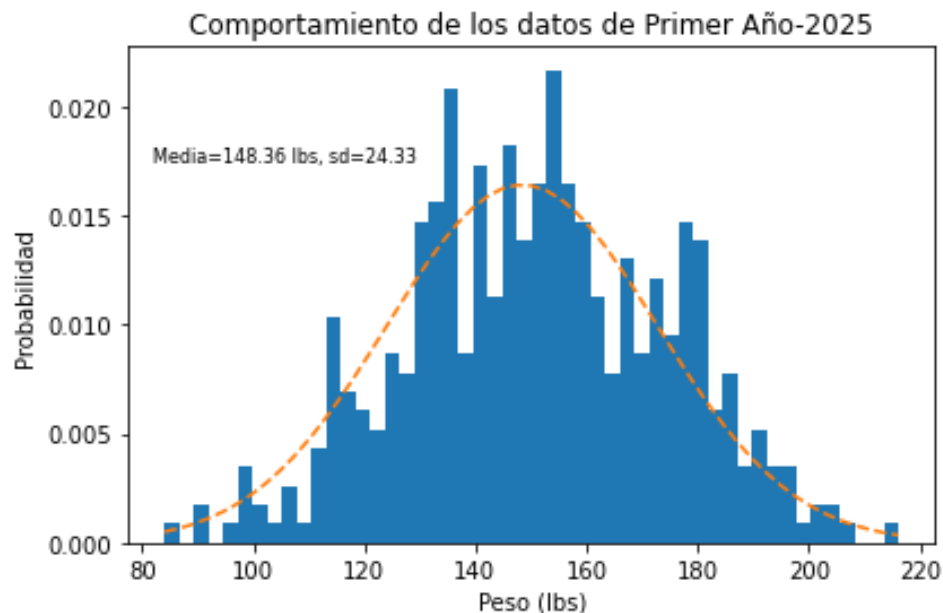


Gráfico 1: Histograma 1er Año

En el siguiente histograma (Grafico 2) es posible observar que los datos intentan encajar en una trayectoria similar a la “Campana de Gauss”, por lo cual se puede determinar mediante esta prueba grafica que en efecto existe una distribución normal en los pesos de la población estudiantil de la clase 2024. Aunado a ello, se obtuvo una desviación estándar de 26.31%, lo cual indica que la mayor parte de los datos de la muestra tienden a estar agrupados cerca de su media equivalente a 147.86 libras.

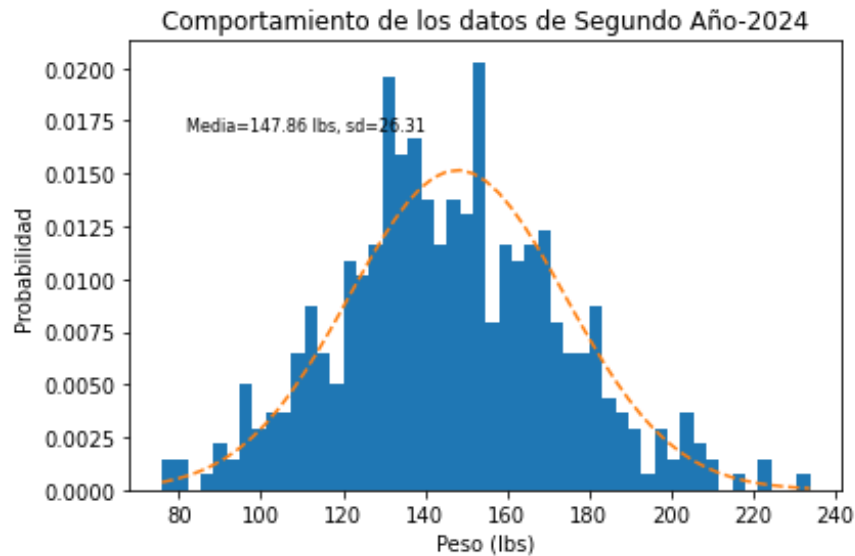


Gráfico 2: Histograma 2do Año

El histograma (Gráfico 3) representa de forma gráfica que los datos buscan la manera encajar en una trayectoria similar a la “Campana de Gauss”, por lo cual se puede determinar que existe una distribución normal en los pesos de la población estudiantil de la clase 2023. Aunado a ello, se obtuvo una desviación estándar de 27.41%, lo cual indica que la mayor parte de los datos de la muestra tienden a estar agrupados cerca de su media equivalente a 159.13 libras.

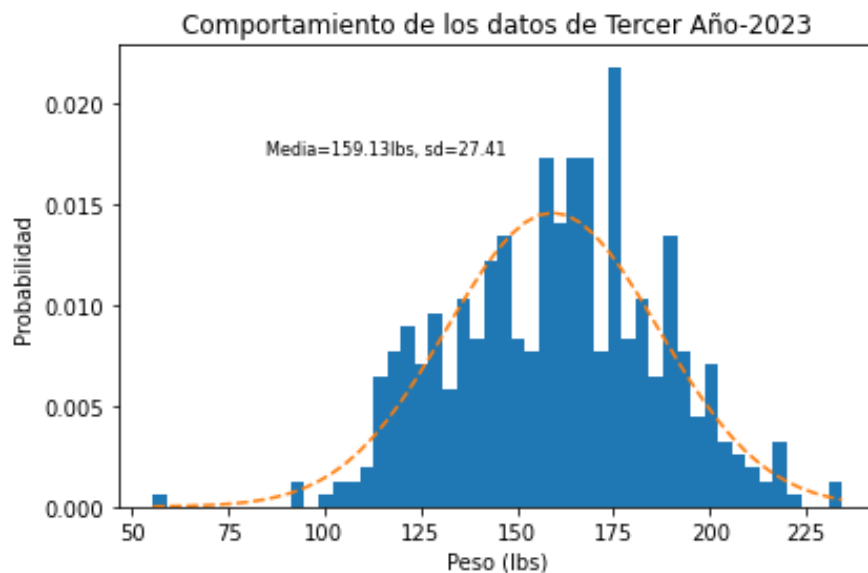


Gráfico 3: Histograma 3er Año

Gráfico de Barras, Peso Promedio por Clase:

El peso promedio de tercero fue de 159 libras, el cual, es mayor en comparación a los otros años, esto podría ser generado porque los estudiantes de este respectivo año tienen cursos que generan que su demanda de alimentos incremente, así pues, su ingesta es superior; asimismo, la actividad física es menor, puesto que, la demanda académica es intensiva, esto ocasiona que los alumnos no dispongan de tiempo libre para realizar actividades que contribuyan en su salud física. (Ver Cuadro 1)

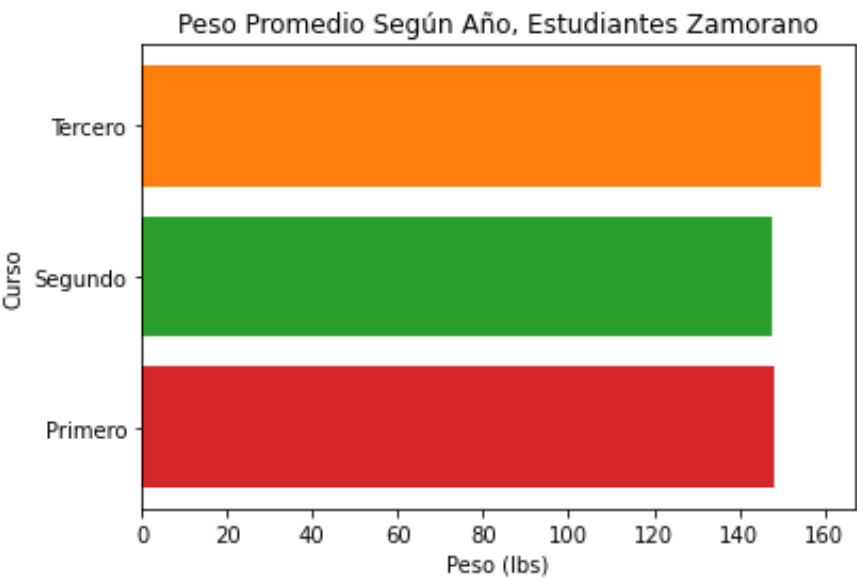


Gráfico 4: Grafico de barras, Peso por Año

Peso Promedio por Clase			
Clase	Primero	Segundo	Tercero
Peso (lbs)	148.37	147.87	159.13

Cuadro 1: Peso Medio 2025, 2024, 2023

Gráfico de Líneas Genero-Peso:

Se puede apreciar que las mujeres presentan un peso por debajo del peso general, esto se puede atribuir a la moda que tienden seguir las femeninas en tener una vida saludable, por lo tanto, su dieta es restringida y realizan ejercicio.

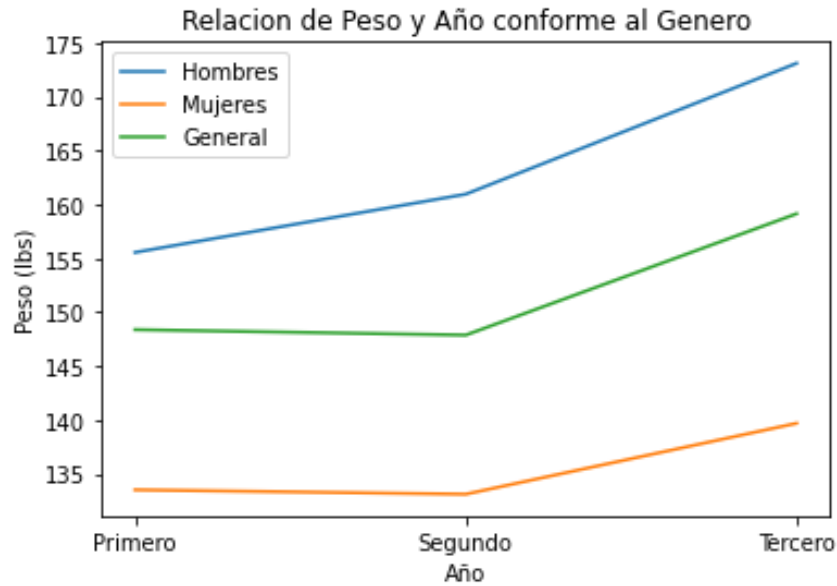


Gráfico 5: Grafico de líneas, comportamiento del peso conforme al año por Genero

Gráfico de Barras apilada Población-estado IMC:

Se observa que la categoría de Bajo Peso en la población en general como en primer año, no sobrepasa el 5%, lo mismo para el resto de las clases, por otro lado, Peso Normal, prevalece abundante con más de 50% para la población en General, y todas las clases, mientras que, sobrepeso, no representa más de la mitad de manera general, y por último obesidad que si se observa un porcentaje marcado. (Ver Cuadro 2)

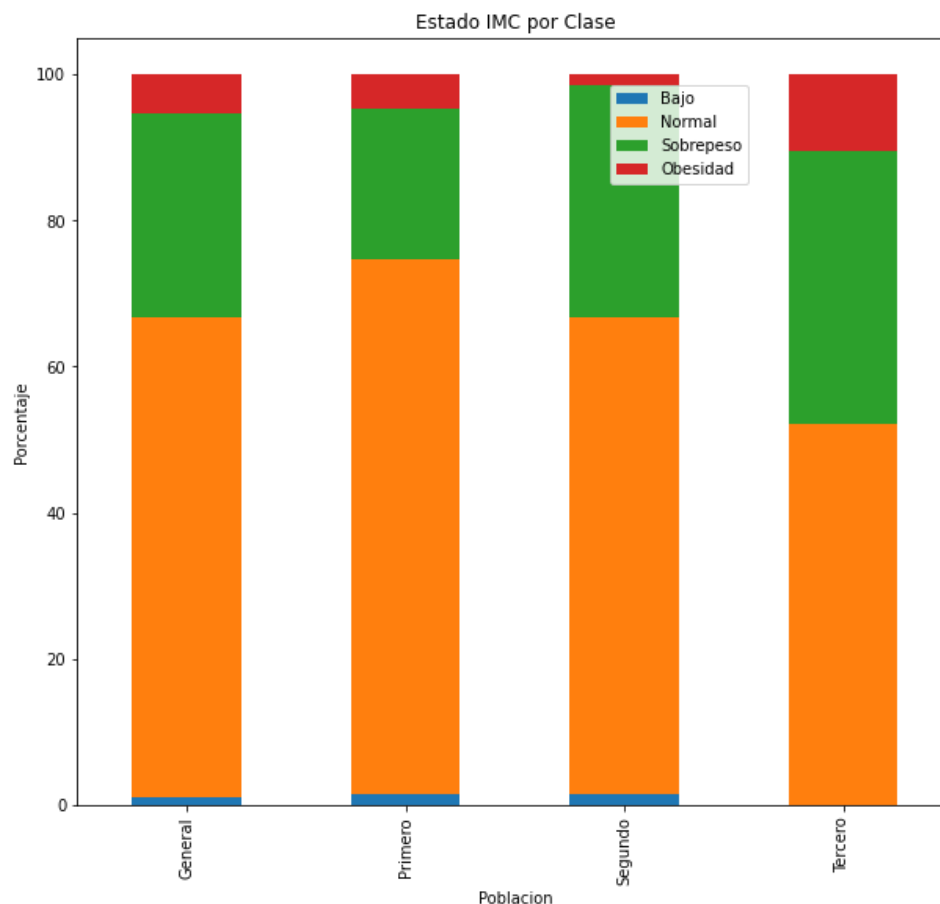
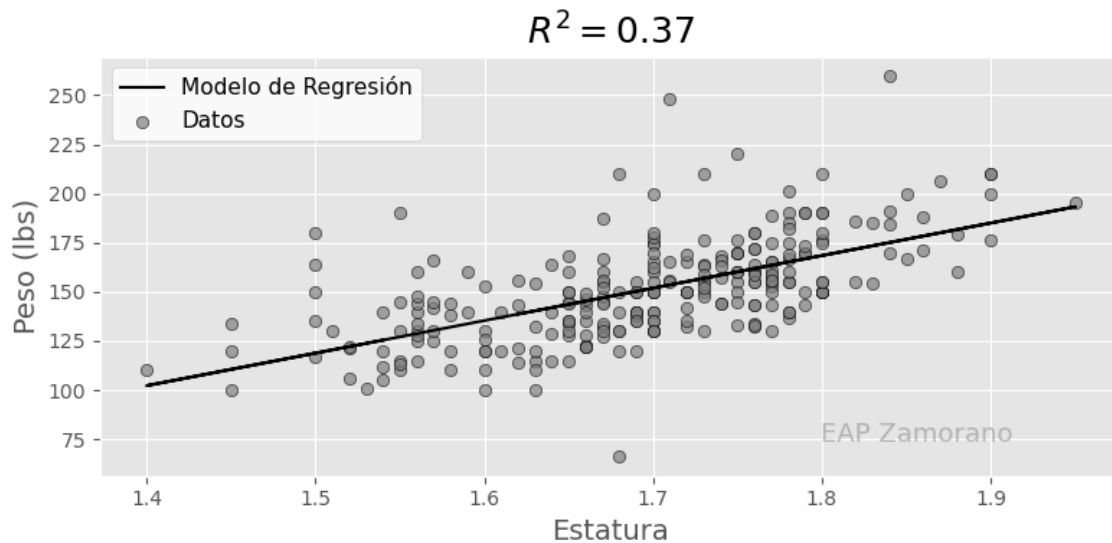


Gráfico 6: Gráfico de Barras apilada Población-estado IMC

Estado de IMC Zamorano (%)				
Año/Estado	General	Primero	Segundo	Tercero
Bajo Peso	1.16	1.00	1.52	0.00
Peso Normal	65.64	73.02	65.16	52.24
Sobrepeso	27.80	20.63	31.82	37.31
Obesidad	5.41	4.76	1.52	10.45

Cuadro 2: Estado IMC Zamorano

Regresión, Correlación lineal de Estatura-Peso y Peso-IMC:



Regresión lineal (Grafico 6) permite predecir el comportamiento de una variable (dependiente o predicha) a partir de otra (independiente o predictora). En este caso la variable dependiente es el peso que se ubica en el eje (Y) y la variable independiente es la estatura que se encuentra en el eje (X) podemos observar que la dispersión de los datos es muy variable respecto a la media, entre la estatura 1.65 m-1.75 presenta una mayor sesgo en los datos, el R cuadrado nos indica que tan alejados están los datos del modelo de regresión, por lo tanto en este grafico el peso y la estatura no muestra una correlación ya que el R cuadrado es de 0.37

Grafico 6: Regresión Lineal, Estatura-Peso

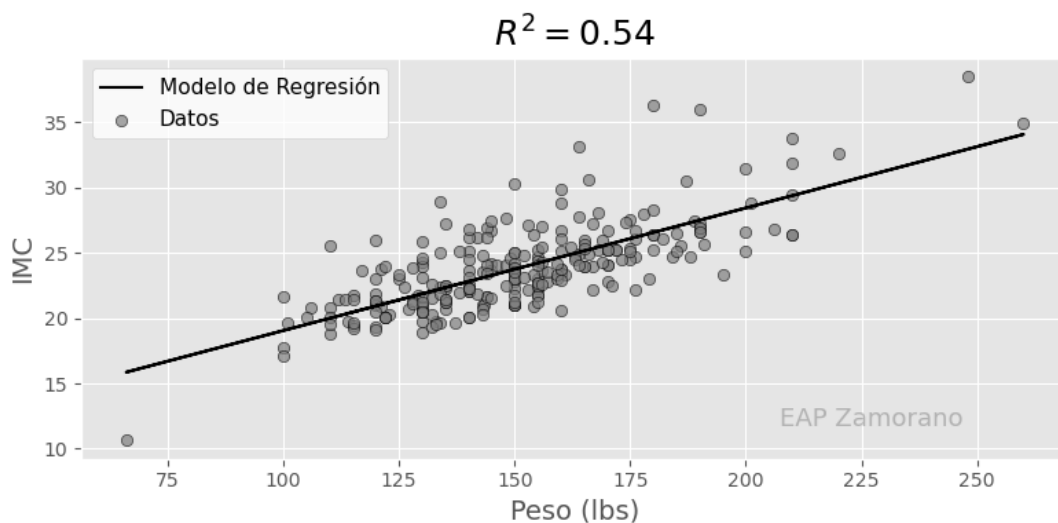


Gráfico 7: Regresión Lineal, Peso-IMC

Regresión lineal (Grafico 7) en este caso la variable dependiente es el IMC (eje Y) y la variable independiente es el peso (eje X). Podemos observar que la gráfica refleja un menor grado de dispersión, en su mayoría los datos se encuentran entre 18-25 de IMC y el R cuadrado es de 0.54 esto quiere decir que es una asociación débil y positiva.

VI. Conclusiones

- Se creo un repositorio en Github el cual sirvio para guardar como respaldo todos los notebooks y datos del proyecto. Asi mismo, se logro elaborar un codigo utilizando la plataforma de Jupyter el cual, permite una facil y rapida digestion y presentacion de datos, en ayuda de librerias externas. El uso de esta plataforma es sumamente util y versatil, para elaborar proyectos tecnico-cientificos de calidad de paper.
- Se observa que la influencia del peso, estatura, y basicamente el año que se cursa en Zamorano, tiene muy poca influencia por sobre el Indice de Masa Corporal de los estudiantes.
- Por otro lado, podemos decir que existe un ligero aumento de peso a partir de tercer año, y realmente muy poca diferencia entre 1er y 2do año. Probablemente, porque en 3er año, existe mayor sedentarismo y estres por el avance de anteproyectos de tesis, cosa que en segundo y 1er año, en lugar de esa actividad, se trabaja arduamente en modulo, lo cual, colabora en la codicion fisica.
- Respecto a la media de la poblacion estudiantil, los varones se encuentran por encima del promedio. Y una gran mayoria de los estudiantes tienen un IMC de Peso Normal, el Bajo Peso, sobrepeso y obesidad, son insignificantes outliers dentro de los estudiantes, pero aun asi son situaciones de las cuales debemos preocuparnos.

VII. Referencias

Bisong, E. (2019). Matplotlib and seaborn. In *Building machine learning and deep learning models on google cloud platform* (pp. 151-165). Apress, Berkeley, CA.

Escuela Militar de Ingeniería (EMI). (s. f.). *Estadística aplicada a la educación superior*.

Recuperado 5 de noviembre de 2022, de

<https://es.slideshare.net/fmartinezsolaris/estadstica-aplicada-a-la-educacin-superior>

Joshi, S. (2021, 30 marzo). *Gráficos de barras apiladas en Matplotlib*. Delft Stack.

<https://www.delftstack.com/es/howto/matplotlib/stack-bar-plots-matplotlib/>

OMS. (2021, 9 junio). Malnutrición. Recuperado 12 de septiembre de 2022, de <https://www.who.int/es/news-room/fact-sheets/detail/malnutrition>

Kim, E. (s. f.). *Multiple Linear Regression and Visualization in Python*. Pythonic Excursions.

Recuperado 5 de noviembre de 2022, de

https://aegis4048.github.io/mutiple_linear_regression_and_visualization_in_python

Reback, J., McKinney, W., Van Den Bossche, J., Augspurger, T., Cloud, P., Klein, A., ... & Seabold, S. (2020). pandas-dev/pandas: Pandas 1.0. 5. *Zenodo*.

Rosales Ricardo, Y. (2012, diciembre). *Antropometría en el diagnóstico de pacientes obesos;*

una revisión. Scielo. https://scielo.isciii.es/scielo.php?script=sci_arttext

statistics — Cálculos estadísticos. (s. f.). Recuperado 5 de noviembre de 2022, de [https://rico-](https://rico-schmidt.name/pymotw-3/statistics/index.html)

[schmidt.name/pymotw-3/statistics/index.html](https://rico-schmidt.name/pymotw-3/statistics/index.html)

Statistics plots — Matplotlib 3.6.2 documentation. (s. f.). Recuperado 5 de noviembre de 2022,

de https://matplotlib.org/stable/plot_types/stats/index.html

Universidad de Alcalá. (2019). *Pandas: herramienta básica para el Data Science*.

<https://www.master data-scientist.com/pandas-herramienta-data-science/>