ADRIAN PINEDA SANCHEZ PORTFOLIO

B.S DATA SCIENCE AND MATHEMATICS ENGINEERING

Adrian Pineda Sanchez

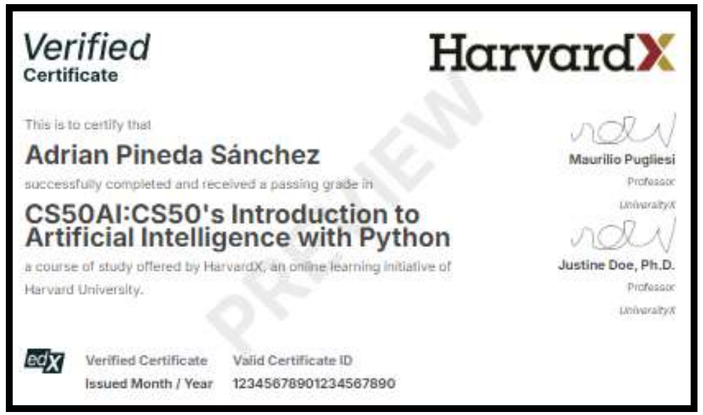
Adrian Pineda

Major in Data Science and Mathematics Engineering
Minor in Advanced Artificial Intelligence for Data Science
Deep Learning Intern Cemex Ventures

Professional Certifications *(Click here to see portfolio Certifications)*

- **Universidad Austral (Coursera)** **Apr 2020 - Jul 2020**
- **Finance: Introduction to Cost Analysis for Business Management**

- **Dataveture in ITESM** **Mar 2023- Mar 2023**
- **Power BI Service**



- **HarvardX by Harvard University (edX)** **Jul 2024 - Present**
- **CS50's Introduction to Artificial Intelligence with Python**

(In progress)

- **Expected completion: Dec 2024**

IS A CERTIFICATE PREVIEW SHOWN BY:edX
Expected completion Dec 2024

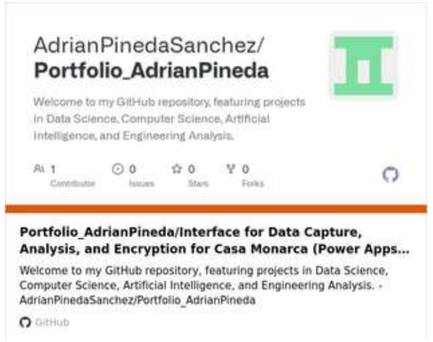
OUTSTANDING PROFESIONAL PROJECTS

[(Click here to see all portfolio Projects)]

Click on the image below to access the Github repository of each project

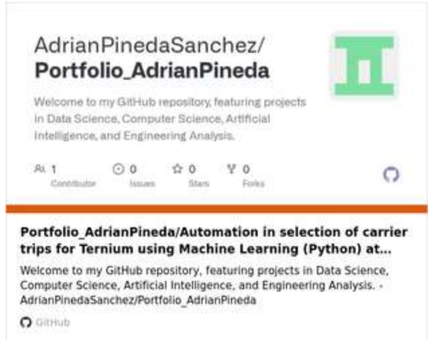
01

Interface for Data Capture, Analysis, and Encryption for Casa Monarca (Power Apps, SQL, Python).



02

Automation in selection of carrier trips for Ternium using Machine Learning (Python)



03

Expense prediction proposal for Hey Banco using machine learning (Python)

04

Creation of graphical interface and data collection for the use of ML in Transportation systems at CEMEX (Python, DAX, SQL)

05

Statistical analysis with SIMA, about concentrations of polluting gases in Nuevo Leon.(R)

Interface for Data Capture, Analysis, and Encryption for Casa Monarca (Power Apps, SQL, Python).

Casa Monarca, a social program in Mexico dedicated to safeguarding migrants, providing them with work, and addressing their needs, required a cryptographic solution to protect migrant registration data due to potential risks. A triple encryption system was designed using asymmetric encryption with AES and SHA-256, along with SSL and TLS protocols for data registration, encryption, and analysis.

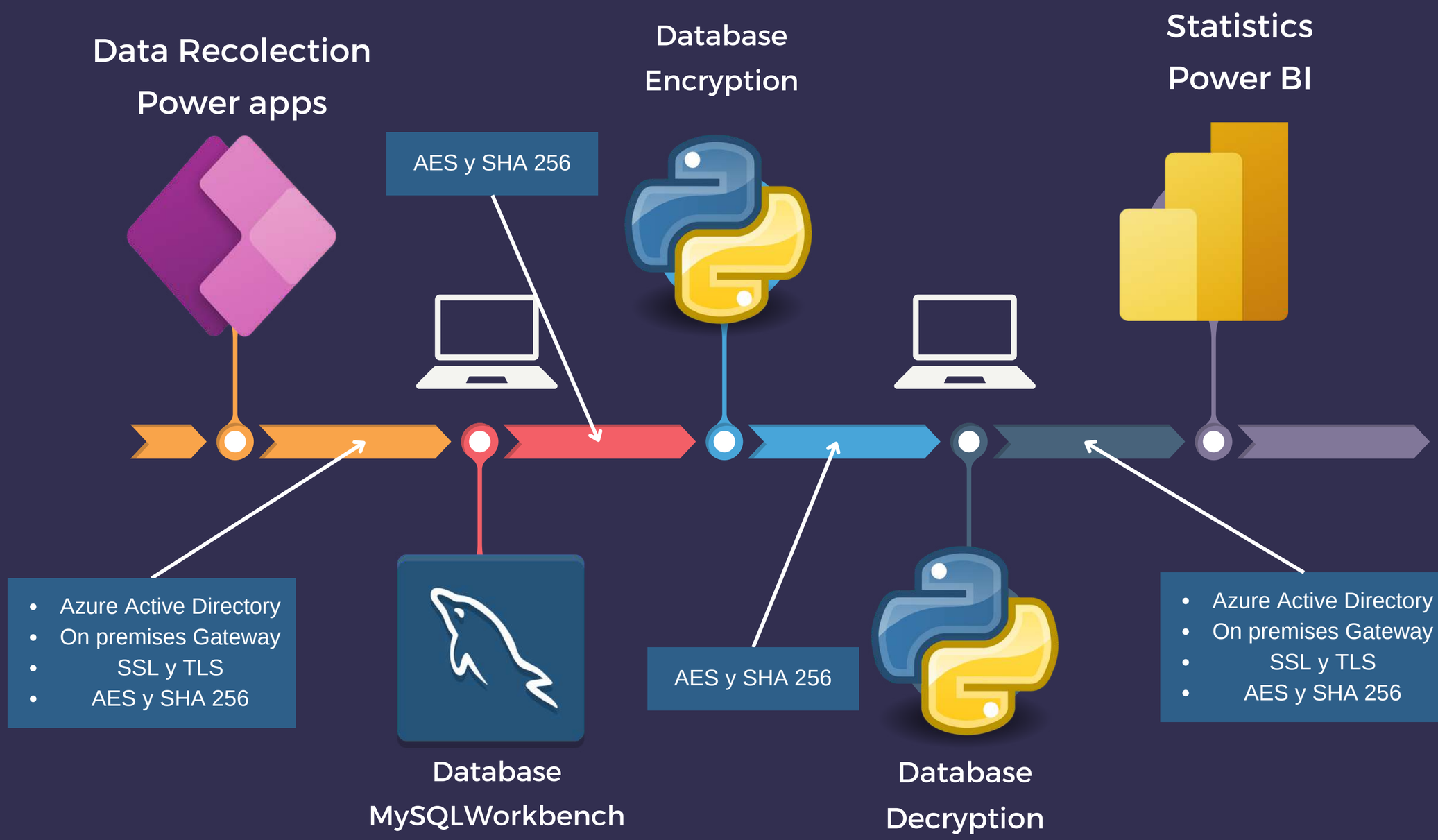
Libraries and Programs

- **Libraries:** Fernet, Numpy, Sklearn
- **Programs:** Power Automate, Power Apps, Power BI, MySQL, Python, LobeAI

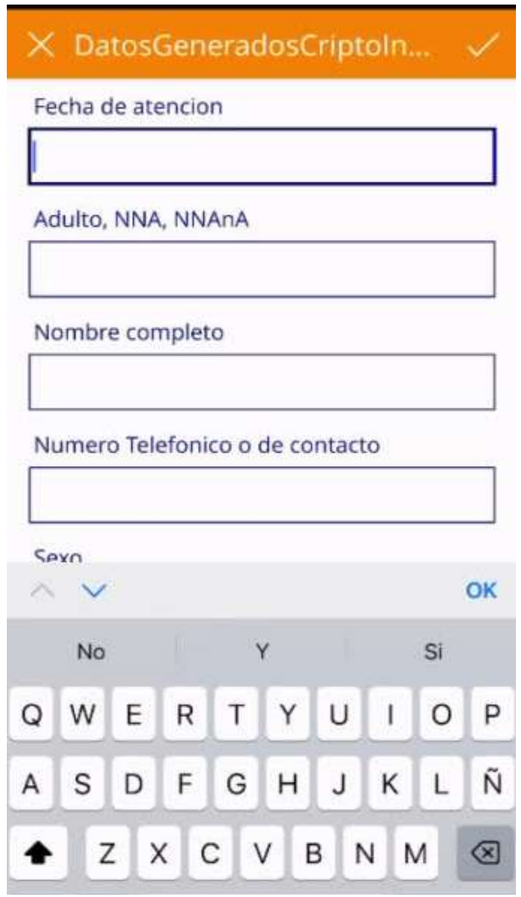
Results

A data collection, encryption, and analysis ecosystem was established for Casa Monarca using free tools, with a potential capacity of 600 records annually.

Flowchart

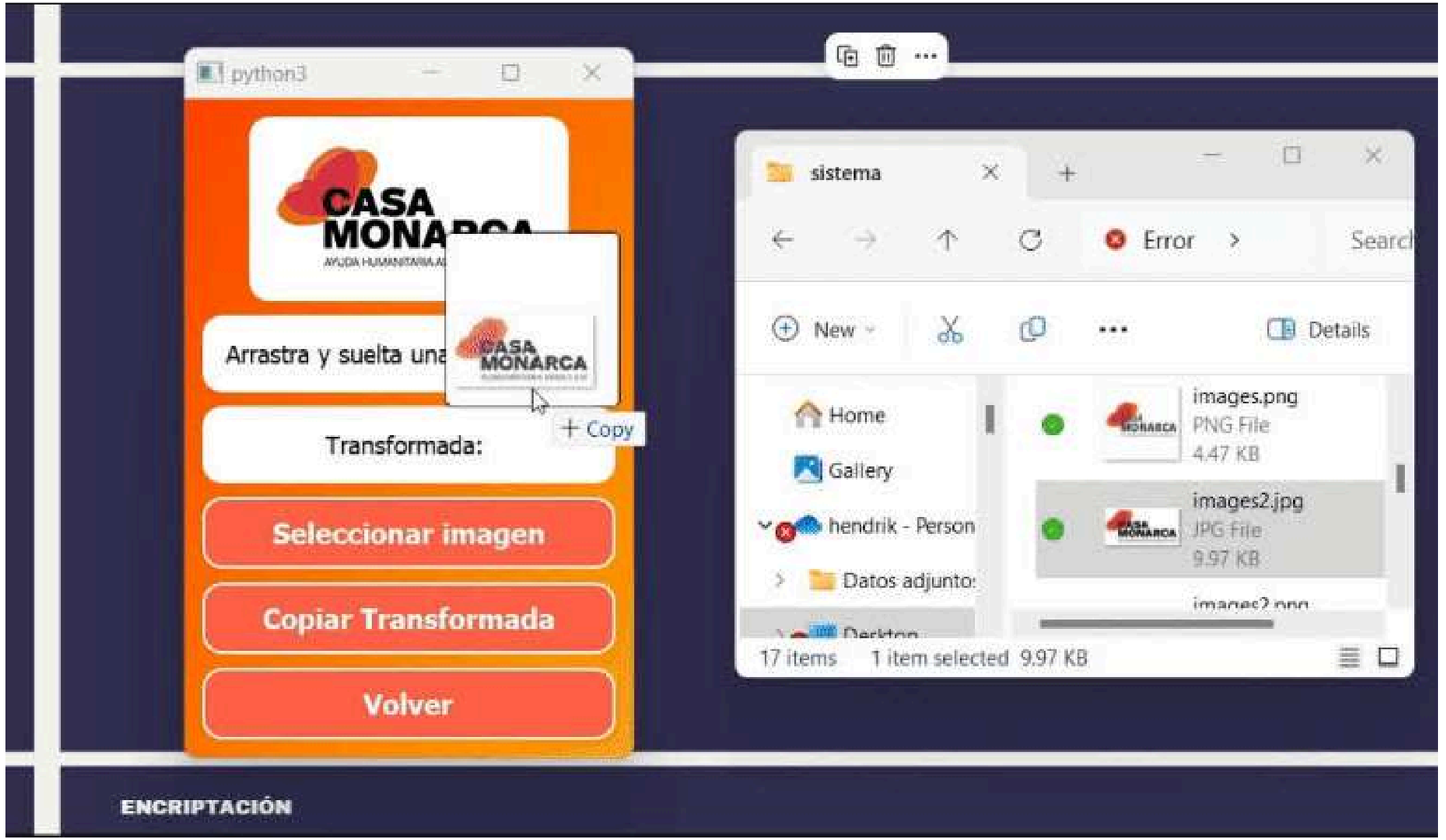

Data Recolection
Power apps
AES y SHA 256
Database
Encryption
Statistics
Power BI
Azure Active Directory
On premises Gateway
SSL y TLS
AES y SHA 256
Database
MySQLWorkbench
AES y SHA 256
Database
Decryption
Azure Active Directory
On premises Gateway
SSL y TLS
AES y SHA 256

Power Apps



Power BI

Local Interface Encryption



Automation in selection of carrier trips for Ternium using Machine Learning (Python)

Through the CRISP-DM methodology, I carried out a process of data exploration, cleaning and transformation, as well as analysis using statistical and data methods, to obtain einsights objectives for the construction of ML models for the prediction of the best driver available. for Ternium, through resource optimization.

Libraries and Models

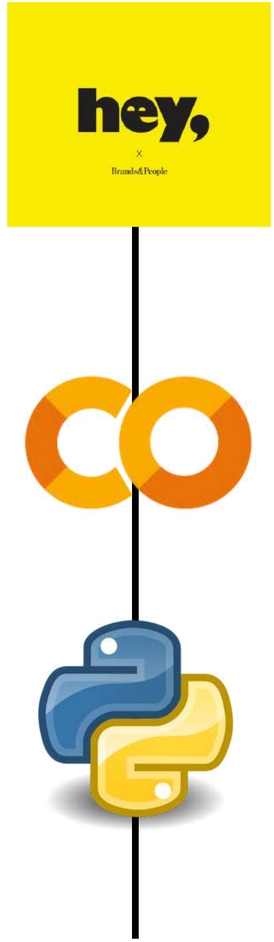
- **Libraries:** Numpy, Matplotlib, Sklearn, pandas
- **Models:** KNN, GBR, Decision Tree, Random Forest
- **Advanced Statistics:** Correlation Matrices, Boxplots Outliers

Results

Through advanced data transformation and statistical methods we were able, through RandomForest, to obtain a Prediction model for Ternium of the most suitable carrier of 90% in the first 3 options

Expense prediction proposal for Hey Banco using machine learning (Python)

Through a database provided in a 24-hour challenge by Hey Banco between university and postgraduate teams, we created value propositions through the search for insights, ending with 3rd place in the competition among more than 85 teams with a personal banking care model through the prediction of expenses and category by client with 97% accuracy



Libraries and Models

- **Libraries:** Numpy, Matplotlib, Sklearn, pandas
- **Models:** KNN, GBR, Decision Tree, Random Forest
- **Advanced Statistics:** Correlation Matrices, Boxplots Outliers, Distributions

Results

3rd place Datathon:
With an Advanced Data Transformation and the verification of several ML models through insights, we were able to obtain a spending prediction model per customer with Random Forest with 97% accuracy

3rd Place

Tecnológico de Monterrey

Por parte del grupo estudiantil **Data Science Club** y

hey, banco

se otorga a:

ADRIAN PINEDA SANCHEZ

el presente reconocimiento por haber obtenido el

3ER LUGAR

durante el evento: **"DATATHON 2023"**

Monterrey, N.L. a 07 de Mayo de 2023.

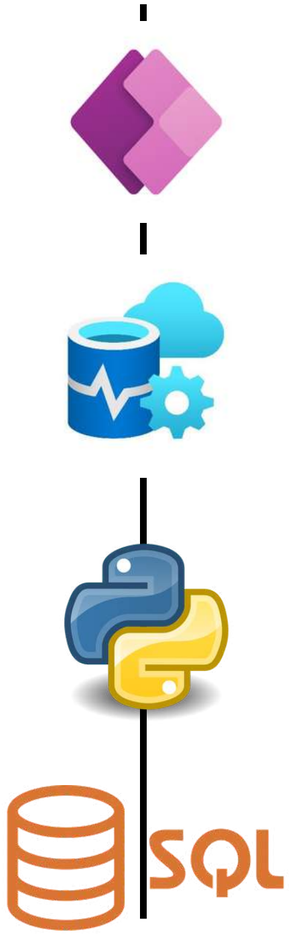
GISELLE SABBAGH
Presidenta

DSC

JORGE DE LEÓN
Vicepresidente

Creation of graphical interface and data collection for the use of ML in Transportation systems at CEMEX (Python, DAX, SQL)

I lead and develop an app project using Power Apps and Power Automate to store in SQL databases in Azure Data Studio, analyzing them with Artificial Intelligence and Machine Learning in Python for insights and predictions on CEMEX's plants and transportation units internationally in a web app.



Libraries and Models

- **Libraries:** Numpy, Matplotlib, Sklearn, pandas
- **Models and Tools:** Random Forest, APIs, Azure Data Studio Encryption, Docker, Power Apps (Automate, Pages, Apps), Dbeaver, etc...
- **Advanced Statistics:** Correlation Matrices, Distributions

Results

- Implementation of **fleet application internationally optimizing logistics** and **transportation management** sectors as well as their plants and resources through **Deep Learning, Power apps user interface, containers** and **Azure databases**

Visualization is confidential

Statistical analysis with SIMA, about concentrations of polluting gases in Nuevo Leon.(R)

Through a compilation of datasets and information explored and mainly by Correlation analysis, PCA, statistical difference tests such as Anderson, Shapiro, ANOVA, etc. we were able to carry out an exploration and find a significant seasonality in the winter months of the rise of particles. PM2.5 and PM10 in sectors of San Pedro

SIMA
Sistema Integral de Monitoreo Ambiental

R Studio®

Libraries and Models

- **Libraries:** Numpy, Matplotlib, Sklearn, pandas
- **Advanced Statistics and models:** Correlation Matrices, Distributions, PCA, Factorial and regression model, discriminant analysis, normality tests

Results

Through discriminant analysis and regression models we were able to find a significant correlation, fulfilling the normality tests between the winter stations and the sudden rise of PM2.5 and PM10 particles in the San Pedro station with the creation of a 94% accuracy prediction model

Simulation test (Cemex Intern)

Predicción de la resistencia de concreto en función de sus propiedades

Descripción
En la industria de la construcción, la resistencia del concreto es una propiedad fundamental para garantizar la calidad y durabilidad de las estructuras. En este problema, se te proporciona un dataset que contiene diversas propiedades del concreto, como la edad, la cantidad de cemento, la cantidad de agua, la cantidad de agregados y más.

Objetivo
Construir un modelo de regresión para predecir la resistencia del concreto con base en estas propiedades (intenta realizar un análisis que destaque respecto a los existentes). Una vez hecho el modelo, crear una aplicación web usando Flask para hacer el deployment del modelo.

Para el assessment, considera utilizar exploración y limpieza de datos donde lo consideres útil.
Para el desarrollo del modelo, considera justificar de alguna manera el tipo de modelo que has elegido.
Para la aplicación web, el requisito es que un usuario pueda hacer uso del modelo introduciendo valores propios de las distintas variables que uses en tu modelo y que la aplicación web muestre la predicción resultante de dichos valores introducidos. Cualquier funcionalidad que quieras adicionar queda a tu consideración y se tomará en cuenta en la evaluación.
Igualmente, el proceso del análisis del dataset es libre, pero se considerará la lógica e insights que resulten de ello para la elaboración del modelo.

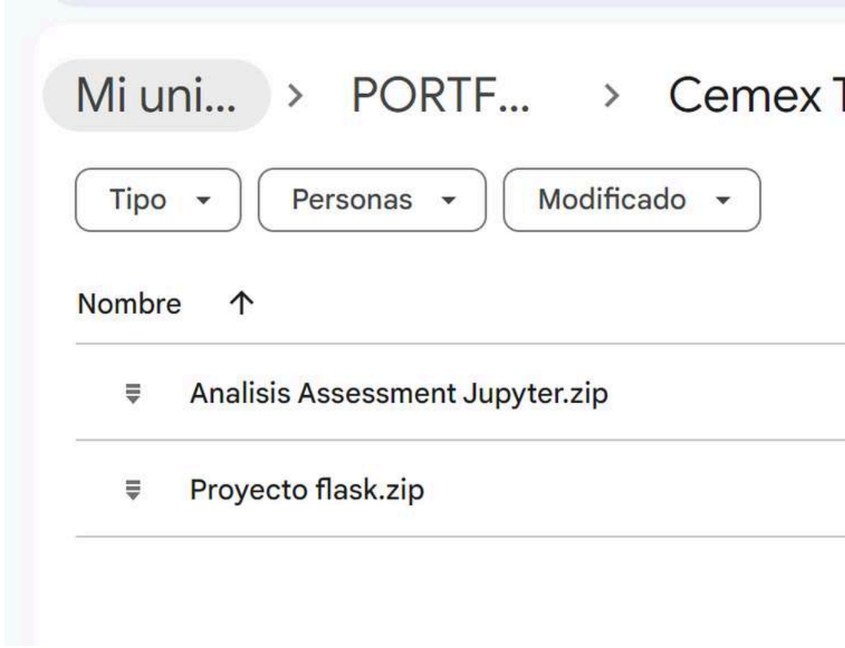
Simulation test (Cemex Intern)

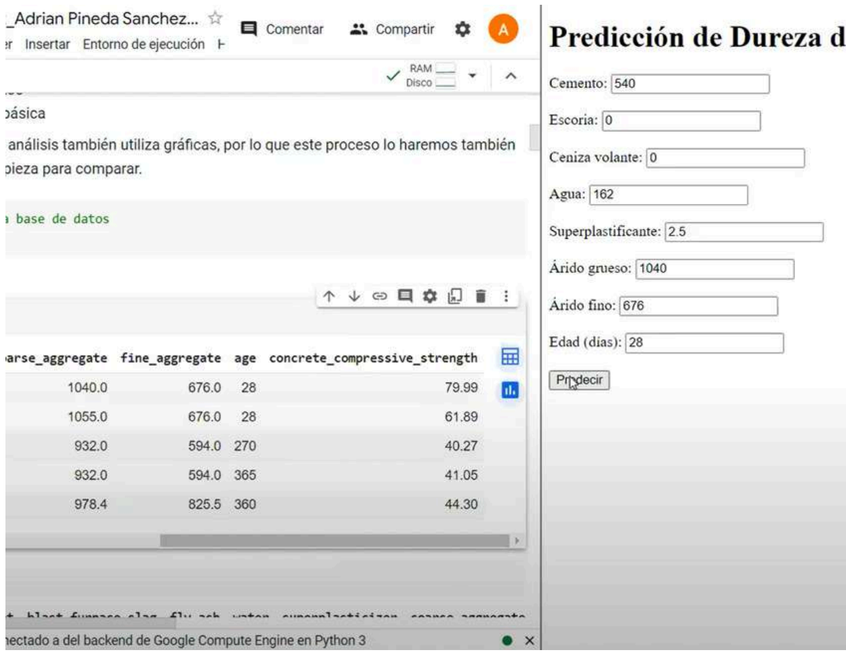
Analysis code in the form of a Jupyter notebook file (.ipynb) (a zip called "Analisis Assessment Jupyter" is included with two deliverables, a PDF of the application of the analysis code with annotations as well as the .ipynb file of the solitary analysis)

The web application in Flask in the form of a folder with the necessary files to run it locally for evaluation. (a zip file called "flask project" is included which contains 3 files in total: an app.py file and a folder called "templates" with 2 .html files necessary for the execution of the web app, "index and result")

link video Example of Execution: https://youtu.be/wTE_Vd594Ao

Files: https://drive.google.com/drive/u/1/folders/1ZqfK2AOZ2hFogrGpn8pWNa9lzSumq6Hg





Thank You