HAMZA AL BUSTANJI DATA SCIENTIST | DATA ANALYST

bustanjihamza@gmail.com 9199319808 Raleigh, NC. https://www.linkedin.com/in/bustanjihamza/ HamzaBustanji

SUMMARY

As a certified Data Scientist with a background in Civil Engineering, I have recently completed two capstone projects which demonstrate my skills in data analysis, data visualization, modeling, and model evaluation. With a passion for learning, a disposition towards people, and a strong work ethic, my goal is to always help those around me and add value to my role to the maximum of my capacity.

EDUCATION

Springboard Aug. 2022 - Aug. 2023
Certificate: Data Science

Philadelphia University of Jordan Jan. 2015 - Dec. 2020
Bachelor's: Civil Engineering

EMPLOYMENT

Springboard, *Springboard Data Science Career Track - Student* Aug. 2022 - Aug. 2023
500+ hours of hands-on course material, with 1:1 industry expert mentor oversight, and completion of 2 in-depth portfolio projects. Mastered skills in Python, SQL, data wrangling, data visualization, hypothesis testing, and machine learning

SKILLS

PYTHON: Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn, Statsmodeles, Streamlit, BeautifulSoup, Selenium, Scrapy
DATA VISUALIZATION: Matplotlib, Tableau, Seaborn
WEB-SCRAPING: Selenium, Scrapy, BeautifulSoup
DATA ANALYSIS:
MACHINE LEARNING: Classification, Regression, Clustering, Feature engineering, Model evaluation
FORECASTING: Time series analysis, Forecasting models

PROJECTS

Favorita Sales Forecast Mar. 2023 - Apr. 2023
Time series analysis and forecasting are critical techniques for businesses and organizations to gain insights into their past performance and make informed decisions about their future. Forecasting, in particular, can help businesses anticipate future trends and make more accurate predictions about demand, sales, and other key metrics. This project aimed to develop a high-level sales forecast for the giant Ecuadorian retailer, Corporación Favorita.
The tools:
-Jupyter Notebook: all work was done using Python written inside Jupyter Notebooks.
-Numpy and Pandas: wrangling, manipulation, summary statistics, and data preparation were done using Pandas and Numpy.
-Matplotlib and Seaborn: data visualization for exploratory data analysis and evaluation purposes was done using Seaborn and Matplotlib.
-Statsmodels: ACF plots, seasonal decomposition, Dickie-Fuller Stationarity Test, monthly plots, quarterly plots, ARIMA model, and SARIMA model.
-Meta's Prophet model.
-Git and GitHub.

House Price Prediction App Feb. 2023 - Mar. 2023
The main goal of this project was to see if we could take a dataset with a large number of features and boil down the features to only a handful that capture most of the information relating to the house price. We then used those features to develop an accessible prediction app using Python's framework Streamlit.
The tools:
-Jupyter Notebook: all work was done using Python written inside Jupyter Notebooks.
-Numpy and Pandas: wrangling, manipulation, summary statistics, and data preparation were done using Pandas and Numpy.
-Matplotlib and Seaborn: data visualization for exploratory data analysis and evaluation purposes was done using Seaborn and Matplotlib.
-Scikit-learn: regression evaluation metrics, dummy regressor for the base model which predicts the mean, GridSearchCV, StandardScaler, TrainTestSplit, LinearRegression, SVR, ElasticNet, RandomForrestRegressor.
-Git and GitHub.

Exploratory Data Analysis of Fandango's Movie Ratings Jan. 2023 - Jan. 2023
Our goal here was to use our data visualization skills in Python to see if we can come to the same conclusion as a FiveThirtyEight article titled "Be Suspicious Of Online Movie Ratings, Especially Fandango's" to determine if Fandango's ratings in 2015 had a bias towards rating movies better to sell more tickets.
The tools:
- Python: Numpy, Pandas, Matplotlib, and Seaborn.
- Jupyter Notebook.

Investigating Car Accidents in the US Feb. 2023 - Feb. 2023
In this project, I went through a dataset containing U.S. car accidents between 2016 and 2021. The dataset was provided by Sobhan Moosavi and has been continuously collected since February 2016 using several data providers.
The goal of this project was to explore the dataset, uncover trends and themes, visualize important aspects of the data in an appealing way, and be able to weave everything into a coherent and engaging story that is understood by everyone, regardless of their technical background.
The tools:
- Python: Numpy, Pandas, Matplotlib, and Seaborn.
- Jupyter Notebook.

Web-Scraping projects
In this repository, I utilize BeautifulSoup, Selenium, and Scrapy to scrape different websites. For example, in the first project, I use BeautifulSoup to scrape the script of the movie Titanic from the website https://subslikescript.com.
The tools:
-Python: BeautifulSoup, Scrapy, Selenium
-Jupyter Notebook