HUDA ALGHAZOULI

404-477-7834 | hudaalghazouli@gmail.com | Atlanta, GA
GitHub: github.com/Hudaalghazouli **| LinkedIn:** www.linkedin.com/in/halghazouli
Portfolio: https://hudaalghazouli.github.io/

A highly motivated Data Scientist with a degree from Georgia Institute Of Technology, boasting a background in data analysis and visualization through group and solo projects. With a bachelor's degree in Applied Mathematics, two years of experience as a Data analyst, one year of experience as a bookkeeper, and three years of experience as a billing specialist, this candidate brings a diverse skill set to the table.

SKILLS

Python, Pandas, Matplotlib, Tableau, Machine Learning, PySpark, SciKit Learn, Hadoop, DataBricks, TensorFlow, AWS, REST API, Flask, SQLAlchemy, SQL, PostgreSQL, Mongo, Excel, VBA Excel, JavaScript, D3JS, Plotly, Leaflet, HTML, CSS, R

PROJECTS

Crystallum Credit Card Company (CCCC) (Python, Pandas, SciKit Learn, SQL, Supervised ML, regression, decision trees)
Data study to predict customer credit approval and the possibility of receiving future payments (group project)
github.com/Hudaalghazouli/Crystallum-Credit-Company
- Merged and analyzed data using Pandas and developed a logistic regression model to predict customer repayment on credit using SciKit Learn.
- Improved model accuracy through multiple regression and classification trials, resulting in an R squared value of 0.99 for testing data and 1 for training data.
- Visualized model performance through ROC curves to showcase the balance between true positive and false positive rates.

Amazon Shopping (Python, Pandas, Big Data, PySpark, ETL, AWS, RDS, SQL)
Performing ETL processes in the cloud for analysis of Amazon's publicly available product review datasets (Solo project)
https://github.com/Hudaalghazouli/Amazon-Shopping-BigData
- Analyzed massive datasets of Amazon's public product reviews using PySpark libraries. Conducted cloud-based ETL processes and utilized DataFrames for further analysis on an RDS instance.
- Created RDS database tables based on the provided schema, and transformed the datasets to fit the tables using PySpark and SQL.
- Loaded transformed DataFrames into RDS, and utilized Python for statistical analysis.

Real-estate Market (Python, Pandas, Matplotlib, Numpy, Gmaps)
Data study of the fluctuating real-estate market in Georgia (group project)
github.com/Hudaalghazouli/Real-estate-Market
- Resolved data generated through Gmaps API to extract coordinates of county seats and generate a heat map based on crime rate data. Used the maps.distance API to calculate the distance from the city of Atlanta.
- Cleansed data with Pandas and utilized Numpy for mathematical operations such as averaging apartment rents.
- Visualized data comparisons through Matplotlib plots and charts, enabling an informed selection of a desirable and secure residential area.

EXPERIENCE

Data Analyst   Oct 2018 – Mar 2020
360 Laboratories   *Lawrenceville, GA*
- Analyzed testing data using Python, Jupyter Notebook, Pandas, and Numpy to identify trends and patterns, providing valuable insights into testing performance and identifying areas for improvement.
- Developed and maintained data analytics and reporting tools using software such as Microsoft Excel, SQL, and Tableau to provide regular reports and dashboards to key stakeholders.
- Developed and implemented data cleaning and processing workflows, resulting in increased efficiency and accuracy of testing processes.

EDUCATION

Data Science and Analytics - Georgia Institute of Technology   Dec 2022
Certificate, Data Science and Analytics.

Bachelor's in Applied Mathematics with a Minor in Information Technology - Georgia Gwinnett College   May 2020