Patrick Salsbury

p.r.salsbury@gmail.com | (408) 887-4634 | LinkedIn | Website | Github

EDUCATION

University of California, San Diego | B.S. Data Science (GPA: 3.5)   *La Jolla, CA / Expected 2024*
- Data Management, Intro to Machine Learning, Applications of Data Science, Data Analysis & Inference, Business Analytics

De Anza College | A.S. Computer Science (GPA: 3.9)   *Cupertino, CA* / 2018 - 2021
- Data Science Fundamentals, Data Structures & Algorithms, Object Oriented Analysis and Design

Certifications
- IBM Data Engineering *(Coursera)* - MySQL, PostgreSQL, IBM DB2, Apache Spark   Present
- Google Data Analytics *(Coursera)* - SQL, Tableau, R   May 2022

TECHNICAL SKILLS

Languages: Python, SQL, Java, Shell Scripting
Databases: IBM DB2, MySQL, PostgreSQL, SQLite3, Redshift
Libraries / Frameworks: Pandas, SkLearn, Apache Spark, PySpark, SciPy, Plotly, Seaborn
Software: Microsoft Azure, AWS, Git Version Control, Databricks, Power BI, Tableau

EXPERIENCE

Computer Science Coach   February 2023 - Present
TheCoderSchool   *Encinitas, CA*
- Coached an average of 5 students per week ranging from Elementary to High School within the topics of computer science.
- Supported the development of apps in **Python** and **Scratch** over a 4-6 week timeframe while simultaneously teaching the fundamentals of coding.

Data Science Intern   June 2022 - September 2022
HM Electronics   *Carlsbad, CA*
- Analyzed customers device data **improving the quality of production and customer relationships** by showcasing areas of concern.
- Led sprint presentations to showcase findings using **Python**, **SQL**, and **Power BI** to influence the decision-making of the product managers and software engineers.
- Collected semi-structured console log data from 100+ customer devices to ingest into a **Azure Blob Data Lake**.
- Automated the ETL process on 30+ million raw records of data at once using **PySpark** on **Databricks(Microsoft Azure)**.
- Predicted customer device failure using **Classification** and **Clustering ML** models to isolate variable indicators for tech support.

Mathematics Tutor   August 2019 - April 2021
De Anza College   *Cupertino, CA*
- Instructed mathematical subjects, such as Precalculus, Calculus, Differential Equations, and Linear Algebra to college students.
- Administered individual study sessions and group workshops, ranging from 5-10 people at once.
- Maximized individual tutee's test results up to the 90%+ range and **increased rates** of passing by more than 50%.

PROJECTS

Popular Youtube Video Titles Generator   Present
- Utilized Youtube APIs to analyze recent top 50 trending videos and predict new possible trending video titles.
- Established an **ETL pipeline** to populate an **SQLite database** after processing words using **Pandas** and **NLTK**.
- Generated possible trending video titles using sentence linguistics, **unigram NLP models**, and **N-gram NLP models**.

Recipe Ratings Analysis and Predictor   February 2023
- Analyzed 234,429 different recipe reviews given on Food.com to **discover underlying associations** and trends regarding rating.
- Concluded statistically significant results using **hypothesis testing** and after performing **exploratory data analysis**.
- Developed a **Decision Tree model** using **CV grid search** after testing models using SKLearn pipelines.

Best Companies to Work For Analysis   July 2021
- Showcased the top 500 best companies to work for in 2021 according to Forbes.com using **Python** and **SQL**.
- Extracted and transformed data using **web-scraping** and **data processing** to store into an **SQLite database**.
- Showcased the top 500 best companies to work for by utilizing **embedded-SQL** queries and **Matplotlib**.