Zaid Patel

Data Analyst

Location: NY | **Phone:** (646) 580-2096 | **Email:** zaid06778@gmail.com | Github | LinkedIn

PROFESSIONAL SUMMARY

Data Analysis and Manipulation:
- Around **3 years** of experience as Data Analyst, handling structured and unstructured data, skillfully overseeing tasks such as Data Mining, Data Acquisition, Data Validation, Predictive Modeling, and Data Visualization.
- Utilized Python packages such as NumPy, Pandas, Matplotlib, Scikit-learn, TensorFlow, and Keras for intricate data manipulation, in-depth analysis, and machine learning.
- Experienced in advanced Time Series Analysis, including ACF plot analysis and GARCH modeling, to enhance data insights for informed decision-making.

Coding Proficiency:
- Employed Python scripts for meticulous data cleansing, incorporating regular expressions, custom functions, and statistical techniques to ensure data accuracy and uphold high data quality standards.
- Demonstrated SQL proficiency for efficient data retrieval, management, and analysis, using complex queries to optimize database performance and ensure data integrity, thereby extracting valuable insights from relational databases for data-informed decision-making.

Machine Learning and Modeling:
- Well-versed in a diverse range of machine learning and statistical algorithms, covering Linear Modeling, Logistic Modeling, Decision Trees, Random Forest, Neural Networks, NLP, Predictive, Descriptive, and Regression Modeling.
- Holds extensive experience in model evaluation, classification, clustering, K-NN, SVM, PCA, and statistical methods.

Data Visualization and Statistical Analysis:
- Excelled in harnessing the capabilities of Tableau and Power BI to develop dynamic and interactive data visualizations, empowering stakeholders to access real-time insights and steer data-informed decisions seamlessly.
- Proficient in leveraging R by utilizing ggplot2 for intricate data visualizations and conducting statistical analysis.

Project management:
- Demonstrated effective project management skills by applying Agile (SCRUM) and Waterfall methodologies to ensure the timely delivery of data-driven solutions.

Cloud Technologies:
- Leveraged cloud technologies like AWS (EC2, S3, Lambda, RDS, DynamoDB, Sagemaker) to scale data analytics solutions, thereby enhancing data processing capabilities and contributing to the efficiency of data-driven projects.

TECHNICAL SKILLS

Programming Languages:	Python, SQL, R (Visualization)
Python Libraries :	Pandas, NumPy, Seaborn, Matplotlib, Transformers, BeautifulSoup, Selenium, Dask
ML Libraries:	Scikit-Learn, TensorFlow, PyTorch, Keras
Databases:	MySQL, Oracle, PostgreSQL, MongoDB
Data Visualization:	Tableau, PowerBI, MS Excel, D3.js
Cloud Technologies:	AWS (S3, EC2, Lambda, RDS, DynamoDB, ElasticSearch, Sagemaker, Redshift)
Tools & Technologies:	Git, Docker, Hadoop, PySpark
Others:	Time Series Analysis, NLP, Data Scraping, Data Wrangling, Feature Engineering, Statistical Analysis

EDUCATION

New York University, Tandon School of Engineering, New York, NY   **Sep 2021-May 2023**
Master of Science in Computer Science

University of Mumbai, Mumbai, India   **Aug 2017-May 2021**
Bachelor of Engineering in Computer Engineering

WORK EXPERIENCE

Data Analyst   **Apr 2023- Current**
Henry Schein   **NY**
- Employed expertise in Python to craft custom scripts for intricate data extraction, transformation, and loading (**ETL**) processes, resulting in a **20%** decrease in labor-intensive manual operations.
- Harnessed capabilities of libraries such as Scikit-learn and TensorFlow, seamlessly integrated advanced machine learning and deep learning models, and achieved a remarkable **15%** improvement in predictive precision for **business forecasting**.
- Demonstrated efficiency and strong foundation in SQL by formulating innovative stored procedures and triggers. These automated mechanisms played a pivotal role in data maintenance, bolstering data reliability, and saving **15%** of time.
- Recognized Power BI's transformative potential, designed and executed ETL workflows, significantly reducing manual data transformation and cleansing procedures, saving **25%** of time and resources.
- Leveraged functionalities of **Power Query** to efficiently connect and transform data from an array of sources, reducing data preparation time by 30% and elevating data accuracy standards.

- Managed **MySQL** databases, fine-tuned table structures, and improved indexing strategies, resulting in a **25%** boost in database query responsiveness.
- **Tools & Technologies:** *Python, SQL, Power BI, Power Query, MySQL, Scikit-learn, TensorFlow.*

Data Analyst Intern   Oct 2022- Mar 2023
Mckinsey   NY
- Conducted meticulous data analysis using Python's libraries and tools to identify key trends, resulting in crucial actionable insights and a notable **15%** surge in overall project efficiency.
- Displayed SQL expertise by optimizing queries, drastically reducing execution times for swift data access and enhanced data-driven analysis speed.
- Leveraged Pandas library to perform complex data cleansing and transformation, addressing multifaceted data discrepancies, and consistently achieving a strong **95%** benchmark in data quality and uniformity.
- Showcased expertise in **Tableau** by creating detailed data visualizations and customized charts to clarify complex data trends, enabling stakeholders to understand underlying patterns and make informed decisions.
- Applied specialized **data mining** techniques to extract invaluable insights from sprawling datasets, leading to an appreciable **10%** uptick in operational processes, streamlining workflows, and maximizing output.
- Incorporated rigorous statistical methods to validate **data hypotheses**, significantly reducing data anomalies by **10%**, enhancing data accuracy and overall integrity.
- Conceived and executed advanced ETL strategies customized for **MongoDB** datasets, overseeing each phase - extraction, transformation, and loading, maintaining impeccable data quality with high accuracy and uniformity.
- **Tools & Technologies:** *Python, SQL, Pandas, Tableau, MongoDB.*

Data Analyst   Nov 2019 - Jul 2021
Adani   India
- Skillfully analyzed intricate datasets using Python, SQL, and R, consistently uncovering at least **3** actionable insights per quarter and translating these findings into strategies driving business growth and innovation.
- Efficiently employed Python libraries, including NumPy, Pandas, and scikit-learn, to conduct in-depth data analysis, resulting in the development of predictive models and data-driven insights that guided business strategies.
- Designed interactive and informative dashboards in Tableau and Excel leading to a **20%** increase in data-driven decision-making across departments.
- Produced comprehensive reports and interactive visualizations from PostgreSQL data, delivering actionable insights to stakeholders and driving a **30%** improvement in data-informed decision-making across departments.
- Skillfully designed and executed robust ETL processes to systematically cleanse, transform, and load data from various sources, achieving over **98%** data quality and consistency through meticulous validation, anomaly detection, and error correction.
- Efficiently harnessed **AWS technologies** like RDS, Lambda, S3, EC2, and Redshift to implement real-time data analytics solutions, resulting in near-instantaneous insights and reducing decision-making time by **30%**.
- Led **Agile (SCRUM)** analytical projects, ensuring on-time and within-scope delivery for **95%** of projects.
- **Tools & Technologies:** *Python, SQL, R, NumPy, Pandas, Scikit-learn, Tableau Excel, PostgreSQL, AWS, Agile (SCRUM)*

PROJECTS

Spark-y Rock Anthem : Song Lyrics Generator - Big Data | Spark | Dask | LSH
- Compiled comprehensive dataset comprising **800,000 song tracks**, totaling **17 GB** in size. The dataset encompasses basic song information scraped from the **Spotify API** and lyrics obtained through the Genius API using **BeautifulSoup** library.
- Implemented **MinHashLSH** search algorithm to facilitate song browsing based on user input of lyrics. Additionally, curated song lyrics generator utilizing **GPT-2** model, generating top-notch lyrics tailored to user preferences.
- **Tools & Technologies:** *Python, BeautifulSoup, MinHashLSH, GPT-2 LLM, NLP*

Movie Recommendation System - Data Mining | Python
- Programmed Recommendation System for movies with **Collaborative filtering** approach and made recommendations via ratings provided by users in **MovieLens 100k** dataset availing **Expectation Maximization Algorithm** with RMSE of 2.51 & **Apriori Algorithm** yielding support greater than or equal to 1.0.
- **Tools & Technologies:** *Python, Pandas*

Real-Time Object Detection - Deep Learning|Android Studio
- Forged android application using Android Studio to detect objects in real-time by executing & optimizing python code to train data using deep learning algorithms. Finalized R-CNN as AUC for **R-CNN & YOLO** stood above **93%**.
- **Tools & Technologies:** *Python, Android Studio, Neural Networks*