**RUSHABH PATEL**

+1 (469) 434-8745 | [patel.rushab@northeastern.edu](mailto:patel.rushab@northeastern.edu) | [linkedin.com/in/rushabh5](https://www.linkedin.com/in/rushabh5) | [github.com/rushabh71](https://github.com/rushabh71) | Boston, MA

**TECHNICAL SKILLS**

|  |  |  |
| --- | --- | --- |
| **Programming & Statistical Languages** | **:** | Python, SQL, R, Java, HTML, CSS, JavaScript |
| **Database Technologies** | **:** | PostgreSQL, SQL Server, Snowflake, MongoDB, InfluxDB, Cassandra, Hive,Visio |
| **Visualization & Analytics Tools** | **:** | Tableau, Power BI, Qlik, MS Excel, Google Analytics, Linux Commands, Git |
| **Python Frameworks & Libraries** | **:** | Pandas, NumPy, Matplotlib, PySpark, Plotly, Sklearn, TensorFlow, Dash, Tkinter |
| **Distributed Computation & ETL Tools**  **Cloud Technologies** | **:**  **:** | Apache Spark, Hadoop, Airflow, Kafka, Talend, SSIS  AWS S3, RDS, Redshift, Athena, EMR, Glue, Lambda, Docker, Azure, GCP |

**PROFESSIONAL EXPERIENCE**

***Nuvera*** ***Billerica, MA • Jan 2022 - Present***

Data Engineering Co-Op

* Developed **ETL pipeline** to extract Engine Test data from IoT devices and export it to **OLAP** DB in the data warehouse
* Orchestrated and monitored daily ETL jobs using **Apache Airflow** to eliminate manual operational overheads
* Enhanced data processing efficiency of BBR **python** **scripts** through code optimization to reduce the runtime by **56%**
* Spearheaded **database optimization** project which resulted in **230%** faster data flow across dashboards, analytical systems
* Designed and deployed a **Dash dashboard** at the organization level to highlight important business KPIs of engine test data
* Worked with engineering and business teams to design and implement **analytics & reporting solutions**

***RHK Automation Solution Ahmedabad, India • Jul 2019 - Jul 2020***

Data Analyst

* Created **Power BI dashboards** to track KPIs and analyze operations & production data to identify new business prospects
* Generated reliable **performance reports** to spot bottlenecks and inefficiencies affecting the supply chain’s efficiency
* Analyzed customer complaint dataset for **descriptive analysis**, drawing out core issues to regain the business loss of **27%**
* Supported cross-functional teams by creating **Ad-hoc analysis**, engaged with engineers, analysts through various channels

**LEADERSHIP EXPERIENCE**

***Northeastern University Boston, MA • Jan 2021 - Present***

Senator | Member of Finance Committee, GSG

* Assisted in creating & maintaining interactive **Tableau dashboards** to keep track of key KPIs for efficient budget allocation
* Represented graduate students’ community in finance committee meetings to ensure budget allocation transparency

**ACADEMIC PROJECTS**

***Sparkify Music Service - Churn Rate Prediction (AWS S3, EMR, Spark MLib, Hadoop, Seaborn)* *Jun 2022 - July 2022***

* Used **PySpark** to perform exploratory data analysis in order to estimate churn using a **12GB** dataset stored in an S3 bucket
* Identified crucial features using **feature engineering** to train **ML models** on the cloud with the highest test accuracy of **84%**

***Olist Retailer - Automated ETL Pipeline (Apache Airflow, Python, Spark, AWS S3)***  ***Feb 2022 - Apr 2022***

* Developed a **Spark job** as a part of an **ETL pipeline** sourcing data from an S3 bucket to track orders that missed deadlines
* Managed data pipeline in **Apache Airflow** to automatically gather necessary insights each day without human participation

***Twitter Sentiment Analysis (Spark Streaming, Kafka, MongoDB, REST API, NLP)***  ***Nov 2021 - Jan 2022***

* Ingested Bitcoin-related tweets data fetched through Twitter API using **Apache Kafka** to perform **Sentiment Analysis**
* Utilized **MongoDB** for publishing output **stream data** with sentiments and Spark for data pre-processing & tokenization

***E-Commerce & Inventory Management Database (SQL Server, MS Visio, SSMS)***  ***May 2021 - Aug 2021***

* Collaborated in a team to design the database conforming to **3NF** with **18** entities to store transactional, inventory data
* Formulated complex **T-SQL queries** & built **server-side modules** to present BI reports, insights to the target audience

***Covid-19 Global Vaccine Distribution Tracking Dashboard (Python, Pandas, NumPy,* Tableau) *Jan 2021 - Feb 2021***

* Cleaned **time-series data** using Python and created **8** dynamic tableau dashboards to track vaccination drive globally
* Analyzed doses to assess national performance together with the accessibility and acceptance of the various vaccines

**EDUCATION**

* **Master of Science in Data Analytics Engineering |** Northeastern University  ***Boston, MA • Expected Dec 2022***
* **Bachelor of Engineering in Mechanical Engineering |** University of Mumbai ***Mumbai, MH, India • Jul 2019***

**CERTIFICATIONS**

* Data Engineering Nanodegree, Udacity (Data Modeling, AWS Cloud Data Warehouses, Data Lake, Spark, Hadoop, Airflow)