

YASH SHIMPI

yashnshimpi@gmail.com | +1 315-603-8308

<https://github.com/yashshimpi29> | <https://www.linkedin.com/in/yashnshimpi> | <https://yashshimpi29.github.io/>

EDUCATION

Syracuse University, School of Information Studies, Syracuse, New York

Master of Science in Information System | GPA - 3.9/4

May 2023

Relevant Coursework: Big Data Analytics, Data Analysis & Decision Making, Data Warehouse

University of Mumbai, Mumbai, India

Bachelor of Engineering in Computer Engineering | GPA - 8.9/10

Oct 2020

Relevant Coursework: Database Management System, Machine Learning, Data Mining

TECHNICAL SKILLS

Certifications: Tableau Desktop Specialist, Microsoft Office Specialist: Excel Associate (Office 2019)

Languages: Python, R, Apache Spark, SQL

Visualization Tools: Tableau, Microsoft Power BI, Google Data Studio

Technologies: Microsoft Excel, Databricks, Google Cloud Platform (GCP), Google BigQuery, Snowflake, Github, JIRA, Postman

WORK EXPERIENCE

Data Analyst, Reorg Research Inc. | New York, NY

Jul 2023 – Present

- Developing over 10 Python scripts aimed at automating daily validation and API-related tasks, resulting in a reduction of manual workload by more than 75%
- Streamlining the procedure of storing over 1,000 rows of data and conducting financial coverage checks by comparing it with an existing dataset of over 100,000 records
- Supporting 2 teams by the creation of ad hoc queries and integrating a new API to fuel product expansion and growth

Research Assistant, Syracuse University | Syracuse, NY

Sep 2021 – Jul 2023

- Formed an ETL pipeline to collect 2.5M+ tweets using Twitter API to understand changes of beliefs in people
- Extracted over 1500 content data by web scraping tools: Selenium and Python and transformed the data items to store in MongoDB for social network analysis between users
- Assisted 5 Fellows in providing statistical and over 20 visual and text insights from Tableau dashboards and NLTK library to support the research for reducing the struggles of female refugees and children

Data Science Intern, RSG Media | New York, NY

May 2022 – Aug 2022

- Developed a Matching Pipeline with an accuracy of 97% to unify data from 2 movie tables, resulting in a thorough database
- Engineered a robust pipeline using Databricks and PySpark to create a database, enabling efficient transformation and loading of 20M+ metadata across 15 tables from the TMDb API to AWS S3 bucket
- Automated the process of sending metadata and metrics as inputs to machine learning models by restructuring a workflow and converting over 10 SQL queries to PySpark format

Data Analyst Intern, Syvylyze Analytics | Pune, India

Sep 2020 – May 2021

- Profiled 900+ store information utilizing Python and Google Maps API to get 8 extra metadata for analyzing retail performance
- Collaborated closely with cross-functional teams to enhance the retail store Data Master in the enterprise data warehouses and effectively communicate data reporting to non-technical stakeholders
- Executed and presented clear documentation of 3 different projects that were an integral part of providing geolocation data for a real-time client project: retail analytics

PROJECTS

Price Analysis and Recommendation for Airbnb

Oct 2022 – Dec 2022

- Conducted analysis of an Airbnb dataset with over 50 columns, providing actionable insights to support growth of Airbnb hosts
- Implemented linear regression and random forest classifier in PySpark to address and accomplish the goal

Recommendation System for Books

Feb 2022 – Mar 2022

- Gathered a dataset of 50,000 books, computed a weighted score, and utilized the sklearn library to calculate cosine similarity
- Leveraged Heroku, Postgres database service and Flask, along with Git for version control, to host and deploy a web application