Yash Shimpi

yshimpi@syr.edu | +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, NY
Master of Science in Information Management (M.S.I.M) May 2023

University of Mumbai, Mumbai, India
Bachelor of Engineering in Computer Engineering October 2020

WORK EXPERIENCE

Data Science Intern, RSG Media | *New York, NY* May 2022 – Aug 2022

- Created an 8-layered Matching Pipeline to match 2 movie databases to get a unified data source with a match accuracy score of 97%.
- Created a production ready pipeline using Databricks and PySpark to perform transformation and loading of 20M+ metadata in 15 tables from TMDB API to AWS S3 bucket.

Research Assistant, Syracuse University | *Syracuse, NY* Sept 2021 – Present

- Extracted 800+ content data using web scraping tools: Selenium and Python and transformed the data items for storing in MongoDB.
- Worked towards creating ETL pipeline for 2.5M+ tweets using Twitter API.
- Created a process for classifying an Event by using Wikipedia API and Machine Learning Model to get an accuracy score of 90%
- Assisting 5+ Fellows in providing statistical and 20+ visual analytics using Tableau to support research and working towards Text Analysis to obtain keywords and insights using nltk library.

Data Analyst Intern, Syvylyze Analytics LLP | *Pune, India* Sept 2020 – May 2021

- Profiled store locations to better analyze retail transactional performance by extracting 900+ store information using Python and Google Maps API to get 8+ additional meta data.
- Performed Extraction, Transformation and Loading (ETL) of raw data to store in PostgreSQL and MySQL for analysis.
- Coordinated closely with 2+ delivery teams in data reporting and augmented the existing Data Master for retail stores within the enterprise data warehouses.
- Executed 3 different projects that were an integral part of providing geolocation data for a real-time client project: retail analytics.

TECHNICAL SKILLS

Programming Languages: Python (scikit-learn, pandas, matplotlib, data processing, pyspark), R, Spark
Tools: Tableau, PowerBI, Excel (Lookup, Pivot), Jupyter Notebook, Cloud Platforms, Heroku, Time Series Forecasting, Hypothesis Testing, A/B Testing, Data Visualization, Version Control (Git/Github)
Databases: SQL (MySQL, PostgreSQL, MSSQL), NoSQL (MongoDB), Microsoft Azure, Databricks, SQL Server

PROJECTS

Book Recommendation – Heroku & Flask [link] March 2022

- Collected 50k book dataset from Kaggle to perform cleaning and manipulation. Calculated weighted score and took top 10% data to calculate cosine similarity. Used Flask for web designing.
- Used Heroku cloud service to host the application and Heroku Postgres database service for storing the data and used Git for pushing the changes to the cloud.

The Office – Meeting Room Booking System [link] Nov 2021 – Dec 2021

- Designed and implemented a database management system for handling the booking system of meeting rooms for a company. Using MSSQL created tables, views, procedure, triggers for working with databases
- Used Microsoft Azure for hosting the database and Microsoft Powerapps for showcasing the different layout of the application.