Restaurant Review Sentiment Analyzer

Cathreen Paul, Diya Kadakia, Thillai Sudhakar

Demo Link: https://drive.google.com/file/d/1uHCdajNHPIHmSrTySxrjUZgI740LcW-X/view?usp=sharing
Github Link: https://github.com/tsudhakar87/restaurant-review-analyzer

PIPELINE GOALS

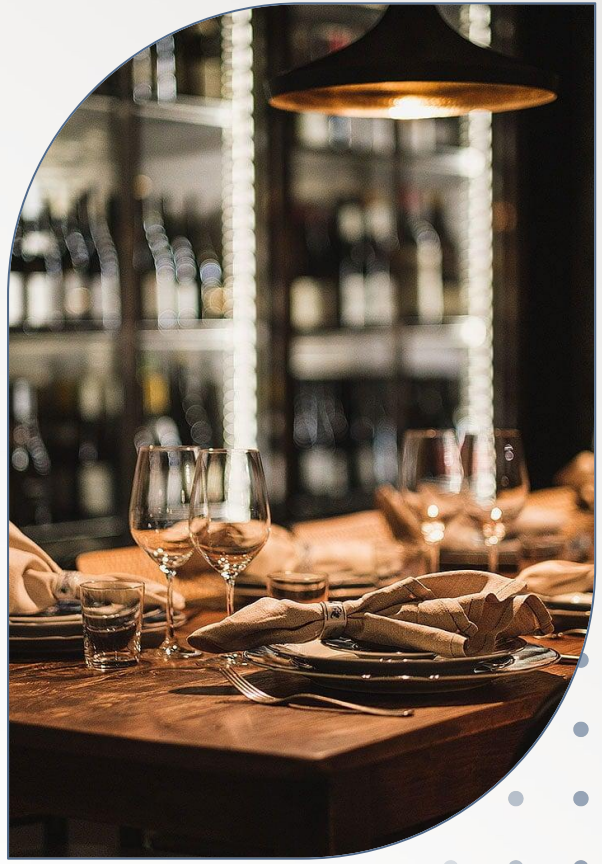
For this project, we wanted to analyze the reviews of restaurants in New York to gain a better understanding of which types of restaurants are popular in the city.

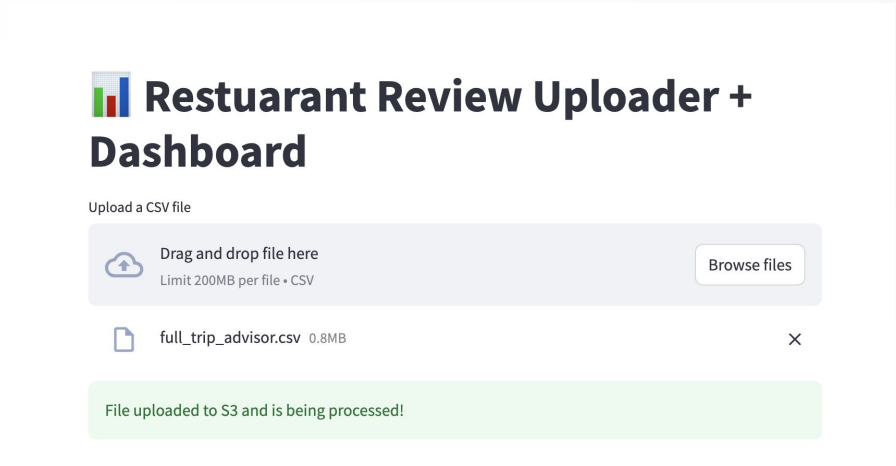
We used a dataset from Kaggle called TripAdvisor New York City Restaurants Dataset, which contained information on the restaurant name, its category/cuisine, the number of reviews, a random review comment, the most popular food, and whether or not online ordering was an option at the restaurant.

Link:
https://www.kaggle.com/datasets/rayhan32/trip-advisor-newyork-city-restaurants-dataset-10k/data

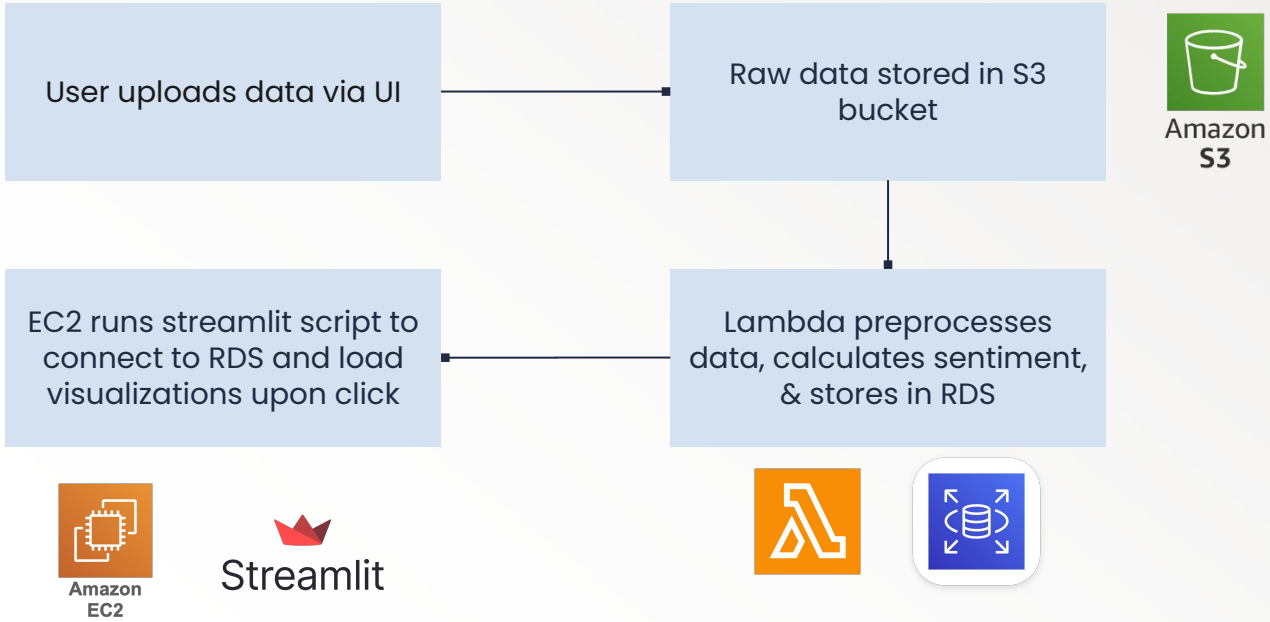
PIPELINE OVERVIEW

Our application provides a UI for a user to upload their restaurant data (the source tripadvisor dataset) and see real time visualizations and analytics.

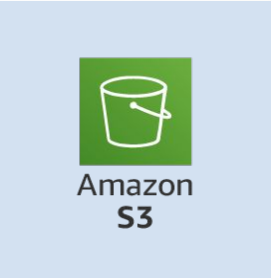




PIPELINE ARCHITECTURE



STORING RAW DATA



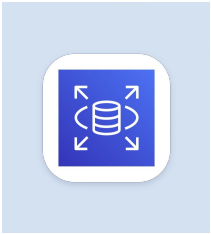
AWS S3

- Created an S3 bucket ‘ds4300reviews-bucket-raw’ to upload and store the csv file
- Object storage optimal for scalability, durability, & security

STORING PREPROCESSED DATA

AWS Lambda

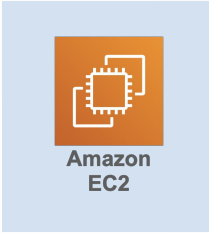
- Wrote functions to clean reviews
- Performed a custom sentiment calculation analysis
- Connected to RDS, tested using uploaded file in S3



AWS Aurora & RDS

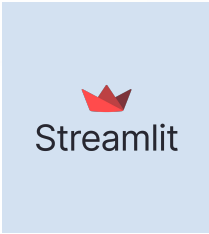
- Lambda handler connected to RDS instance
- Made a database using MySQL
- Inserted data into tables using SQL query

DISPLAYING ANALYTICS VIA EC2



AWS EC2

- Ubuntu Linux AMI, 8 GiB Storage
- Created a local environment to SSH into EC2 instance using necessary credentials
- Ran streamlit script and used external url for local testing



Streamlit

- Built UI where user can upload a CSV file of restaurant review data
- After upload, user can see charts displaying sentiment distributions for different categories & popular foods

TOPICS/SKILLS RESEARCHED

Finding Datasets

We needed to find a dataset that was relevant to our project and suitable for preprocessing. We also wanted a dataset with diverse restaurant categories.

Sentiment Analysis

We had to figure out how to perform sentiment analysis without using python libraries that would not be easily compatible with lambda on AWS. Ultimately, we made our own scoring system by assessing the frequency of positive and negative words in the review.

AWS Services

We hadn't used AWS services like IAM, S3, RDS, Lambda, and EC2 before, so it was definitely a little bit of a challenge understanding how to use these tools. The console helped navigating these easier, though.

DEMO TIME

See everything in action!