CUISINE COMPASS

A restaurant recommendation system using sentiment analysis of user reviews
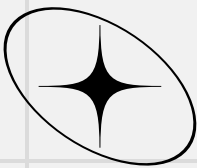
REPORT

LOHITH SRIKANTH PENTAPALLI
RITIKA RADHAKRISHNAN
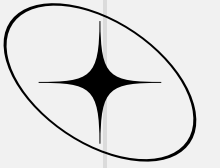LISHEL AQUINAS
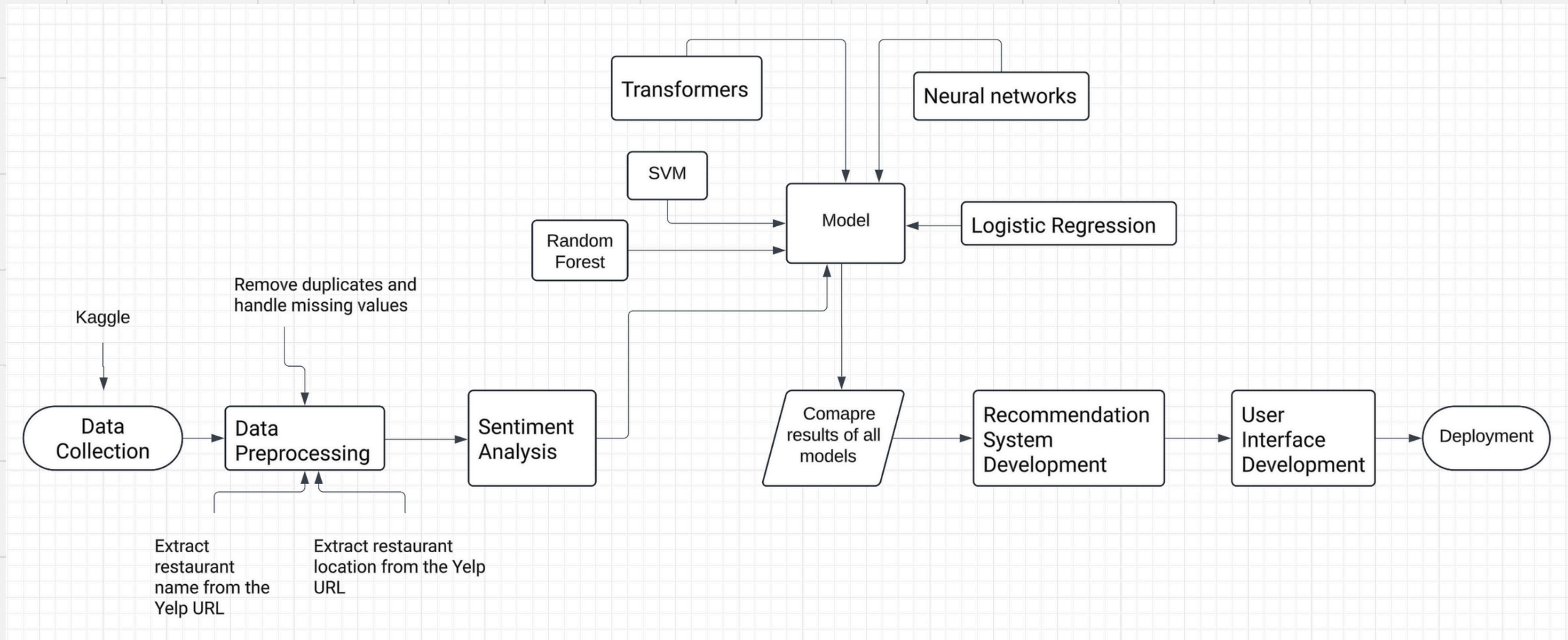DAYASAAGAR
RELINA VAS

TABLE OF CONTENTS

OVERVIEW

Cuisine Compass - A smart restaurant recommendation system that analyzes user reviews using advanced NLP techniques

PERSONALIZED RESTAURANT SUGGESTIONS BASED ON SENTIMENT ANALYSIS, TOPIC MODELING, AND RATING AGGREGATION

RECOMMENDATIONS BASED ON FOOD RECOMMENDATIONS EXTRACTED FROM POSITIVE REVIEWS USING NAMED ENTITY RECOGNITION (NER) AND LOCATION-BASED SUGGESTIONS

COMPARATIVE ANALYSIS OF SENTIMENT ANALYSIS MODELS AND A USER-FRIENDLY INTERFACE FOR SEAMLESS INTERACTION

FLOWCHART

Transformers

Neural networks

SVM

Model

Logistic Regression

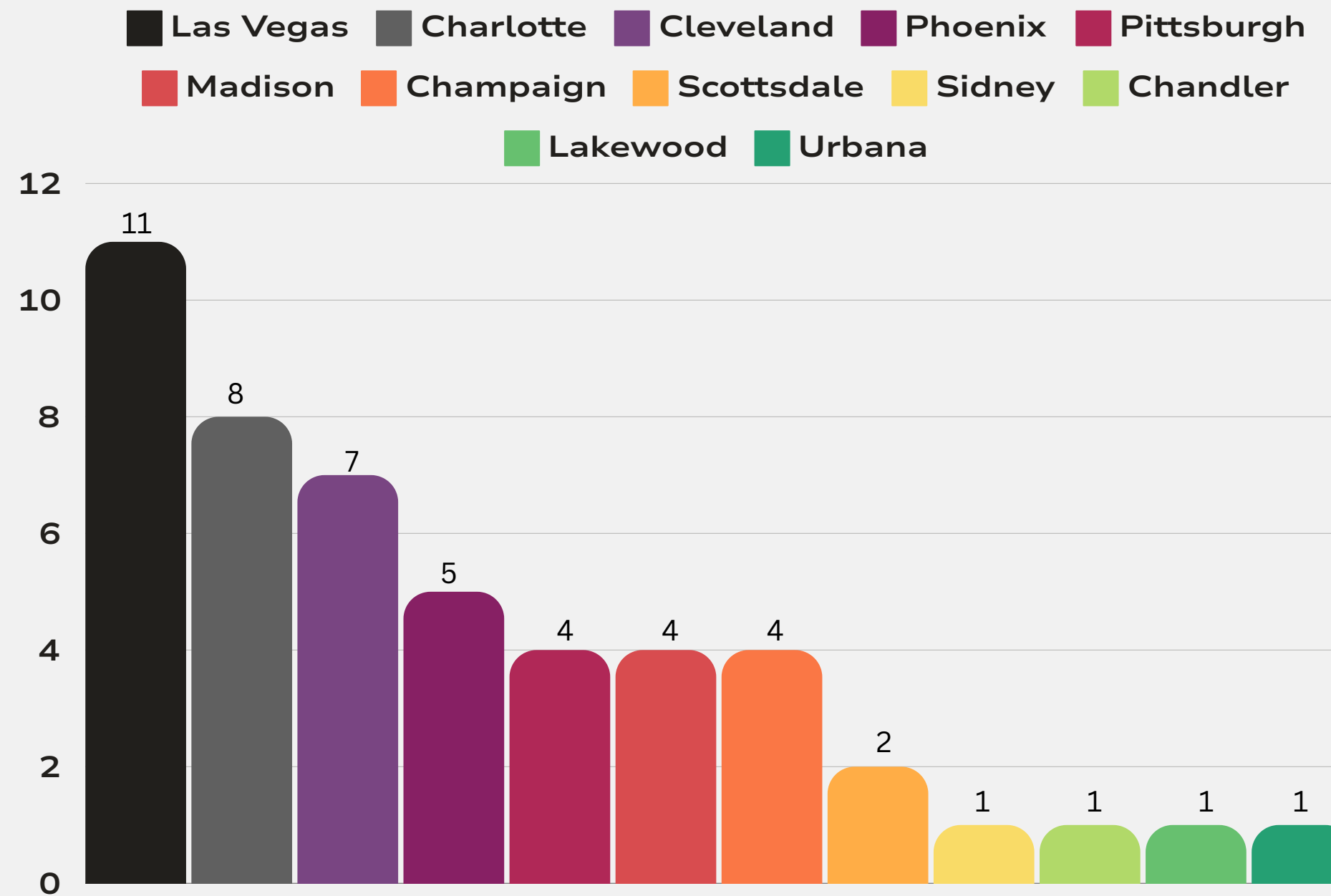Random Forest

Kaggle

Remove duplicates and handle missing values

Data Collection → Data Preprocessing → Sentiment Analysis → Comapre results of all models → Recommendation System Development → User Interface Development → Deployment

Extract restaurant name from the Yelp URL

Extract restaurant location from the Yelp URL

DATA ANALYSIS AND VISUALIZATION

COUNT OF RESTAURANTS IN EACH CITY

Legend:
- Las Vegas
- Charlotte
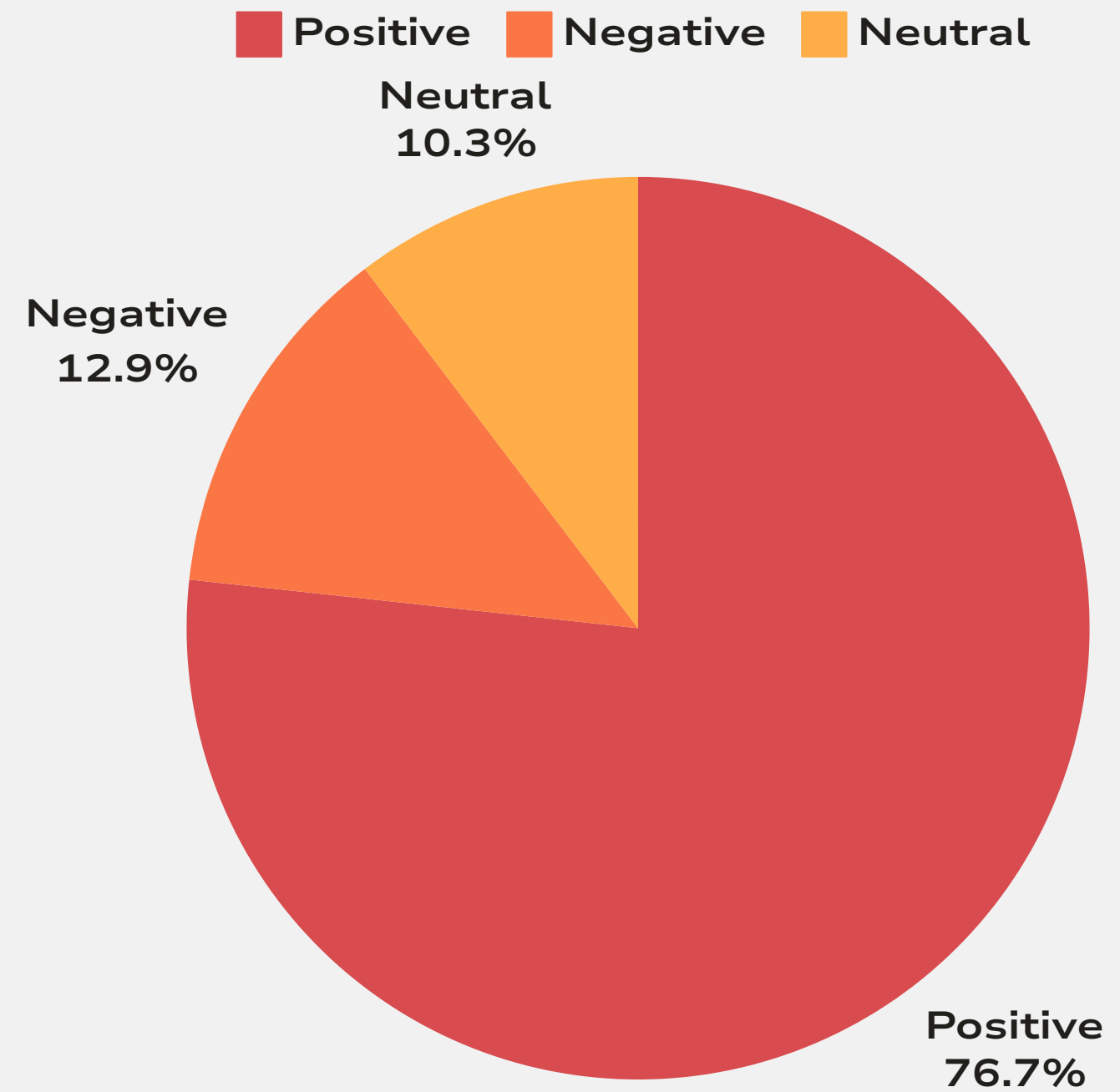- Cleveland
- Phoenix
- Pittsburgh
- Madison
- Champaign
- Scottsdale
- Sidney
- Chandler
- Lakewood
- Urbana

City	Count
Las Vegas	11
Charlotte	8
Cleveland	7
Phoenix	5
Pittsburgh	4
Madison	4
Champaign	4
Scottsdale	2
Sidney	1
Chandler	1
Lakewood	1
Urbana	1
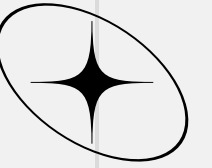
COUNT OF INDIVIDUAL SENTIMENTS

Positive | Negative | Neutral

Neutral
10.3%

Negative
12.9%

Positive
76.7%
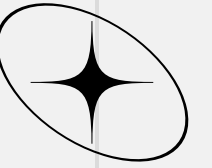
AVERAGE RATING BY LOCATION

Legend: Las Vegas, Charlotte, Cleveland, Phoenix, Pittsburgh, Madison, Champaign, Scottsdale, Sidney, Chandler, Lakewood, Urbana

Values: 4.16, 4.33, 4.25, 4.42, 4.80, 3.78, 4.45, 4.55, 4.39, 4.28, 4.42, 3.11
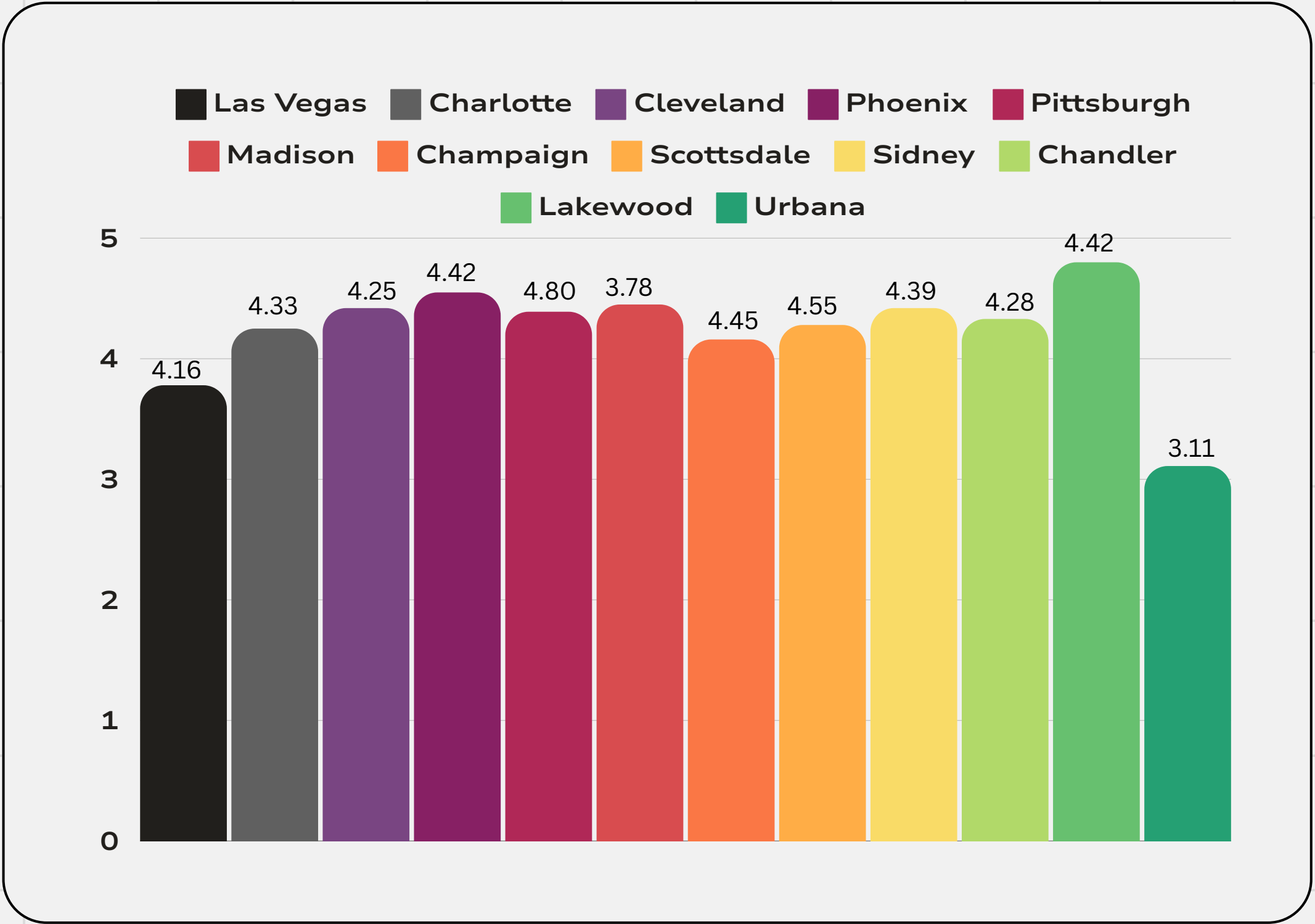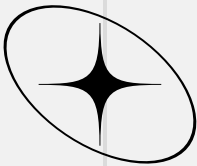
EDA
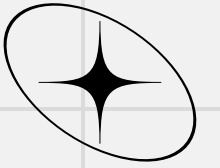
- Data preprocessing involved cleaning URLs, extracting location and establishment names, and reordering columns

- Numerical ratings were transformed into categorical sentiment labels to aid analysis
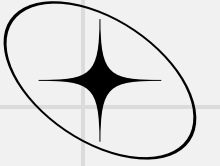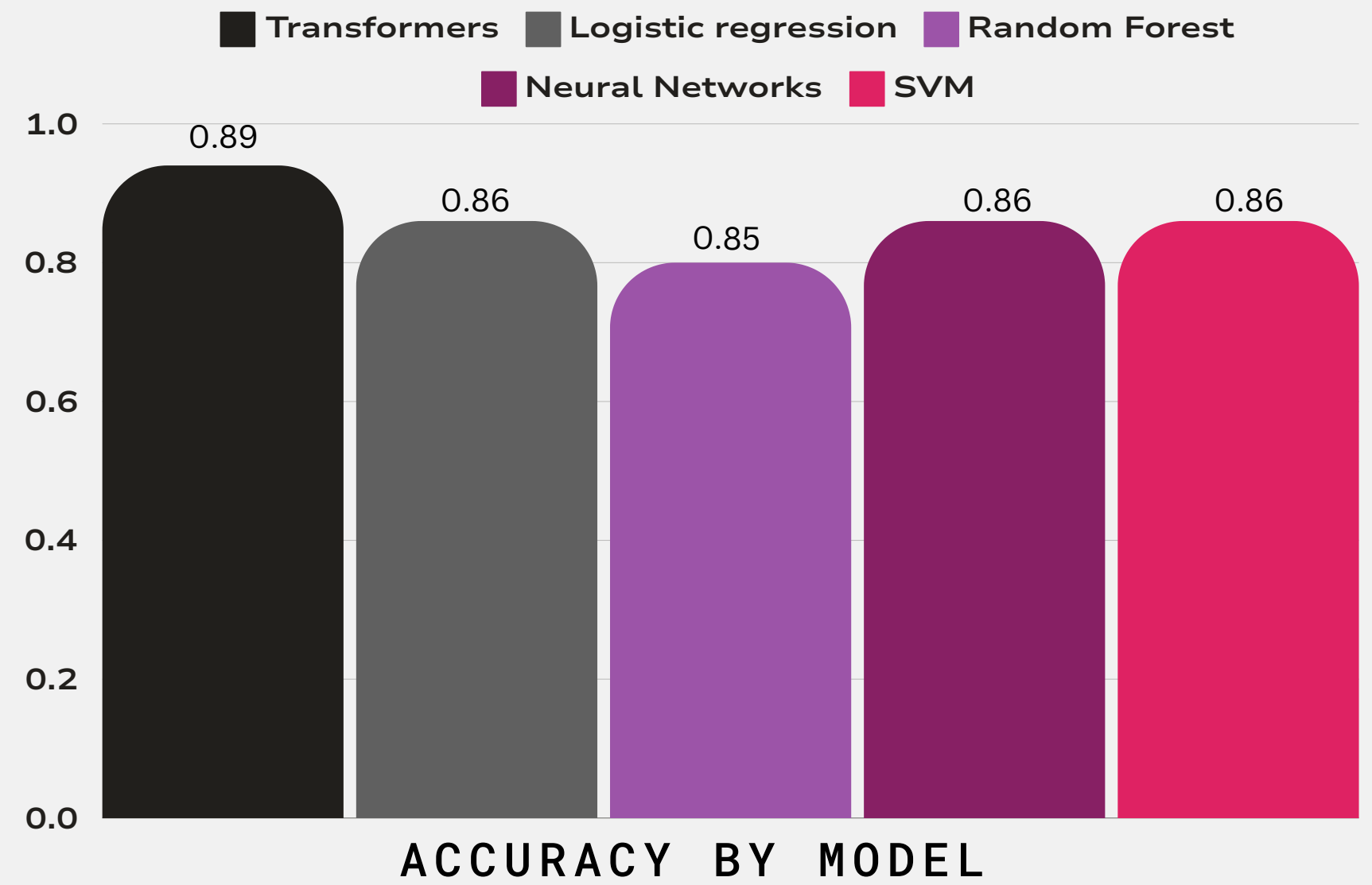
- Exploratory data analysis (EDA) covered the distribution of ratings across locations

- EDA also focused on the frequency of positive, neutral, and negative reviews

- Visualizations such as bar plots, word clouds, and sentiment histograms were used to illustrate insights

SENTIMENT ANALYSIS

- Transformers: Leverage advanced pre-trained models like BERT that capture the context and nuances of textual data in reviews which led to superior sentiment classification accuracy
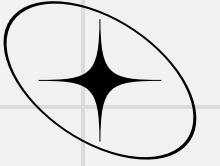
- Logistic Regression: Straightforward and computationally efficient way to model the probabilities of different sentiment classes based on TF-IDF features

- Random Forest: Handle the high dimensionality and sparse nature of TF-IDF vectors from review texts, robustness against overfitting through its ensemble approach

- Neural Networks: Ability to learn complex patterns through layers and non-linear transformations, utilize dropout and batch normalization to enhance generalization

- SVM (Support Vector Machine): Focus on constructing hyperplanes in a transformed feature space to separate different sentiment classes

RESULTS

F1 SCORE BY MODEL

Legend: Transformers, Logistic regression, Random Forest, Neural Networks, SVM

- Transformers: 0.89
- Logistic regression: 0.85
- Random Forest: 0.74
- Neural Networks: 0.84
- SVM: 0.85

ACCURACY BY MODEL

Legend: Transformers, Logistic regression, Random Forest, Neural Networks, SVM

- Transformers: 0.89
- Logistic regression: 0.86
- Random Forest: 0.85
- Neural Networks: 0.86
- SVM: 0.86
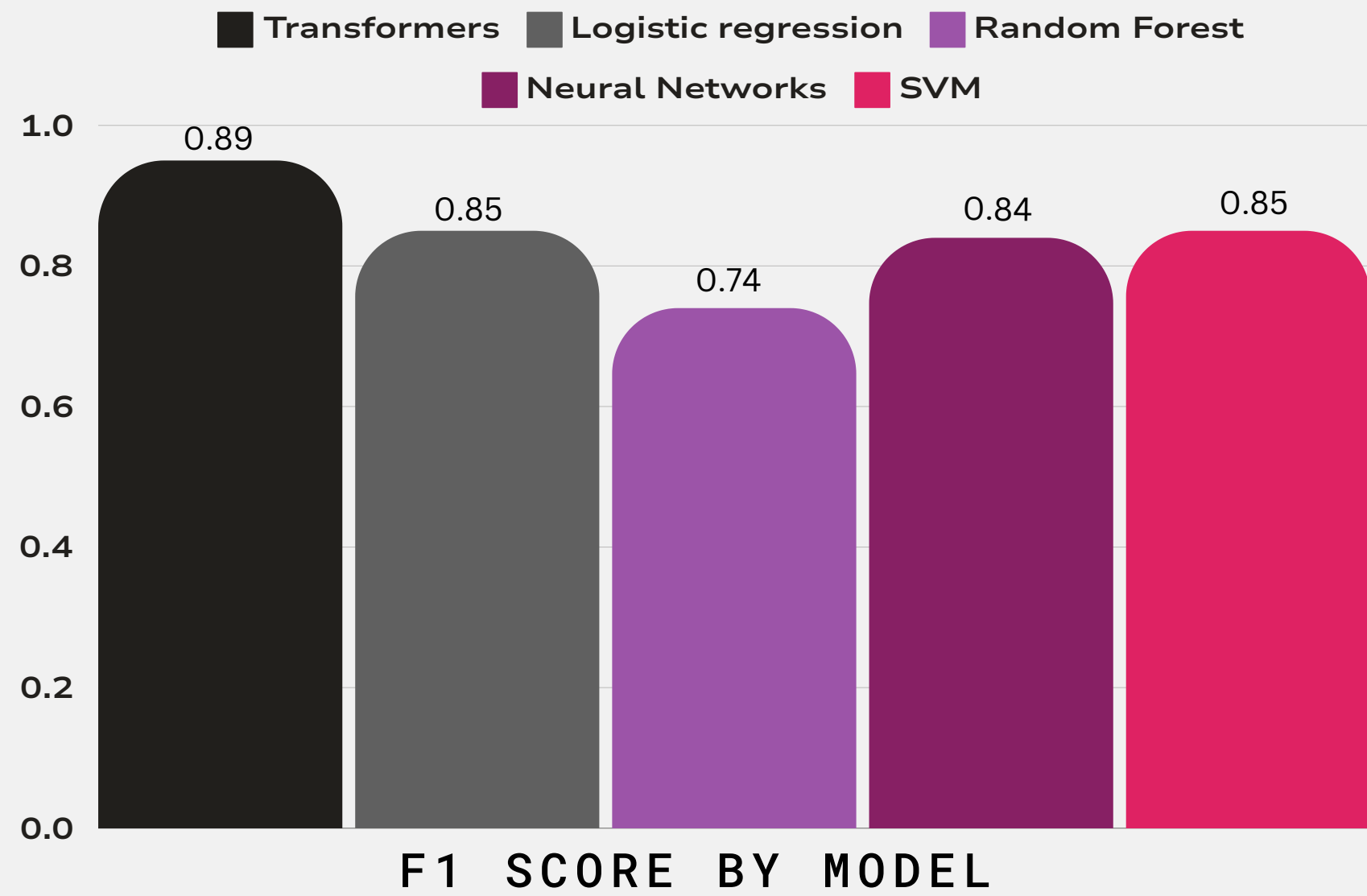
RECOMMENDATION SYSTEM

- The SentimentIntensityAnalyzer from the NLTK Vader module offers a straightforward approach to sentiment analysis in text data

- This tool evaluates the intensity of positive, negative, and neutral sentiments, along with an overall compound sentiment score

- By assigning polarity scores to individual tokens (words or phrases), it aggregates them to generate an overall sentiment score ranging from -1 to +1

- Its versatility allows it to handle informal text found in social media, including slang, emojis, and context-specific expressions

- The SentimentIntensityAnalyzer is invaluable for applications such as social media monitoring, customer feedback analysis, and opinion mining
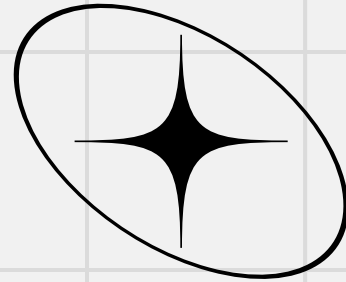
RESULTS
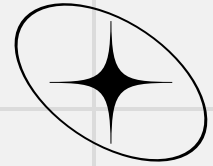
- Food Recommendations: Using Named Entity Recognition (NER), curated recommendations for must-try dishes derived from positive reviews

- User engagement: Catering to individual tastes and preferences, fostering culinary discovery and enjoyment

- Location-Based Recommendations: Recommend nearby dining establishments tailored to their unique preferences and interests using user location data

- User Interface: Using Streamlit, an open-source Python library

- Display the output that includes the top restaurants in the given location along with some must try dishes from those restaurants

- User drop down in the UI will have recommendations that have all restaurants from highest rating to the lowest, must try recommendation in each restaurant, and top 5 reviews for every restaurant

FUTURE SCOPE
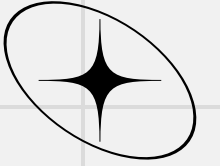
- Dynamically adjust restaurant suggestions based on the time of day and user preferences, E.g. brunch during weekends and fine dining options for dinner on weekdays

- Highly personalized suggestions with integration of user-specific data, such as dietary preferences, past dining experiences, or favorite cuisines

- Real-time feedback mechanism to continuously refine and improve its recommendations

- Make recommendations relevant and personalized over time

THANK YOU