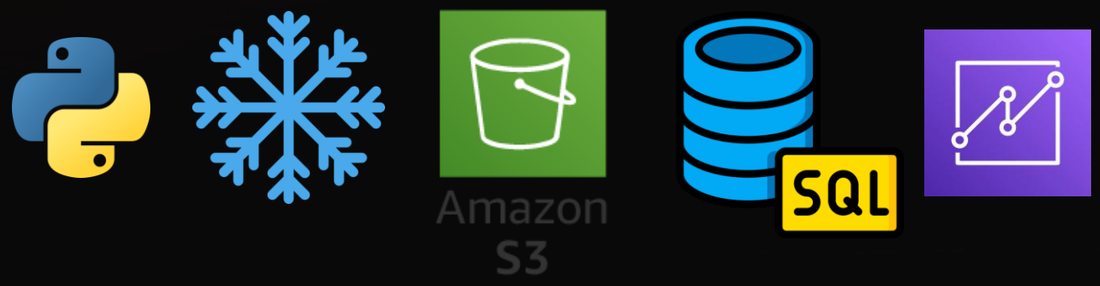
yelp Data Analysis

"Comprehensive analysis of 7 million Yelp records across 114k businesses and 150k reviews"



Data Pipeline Architecture

Yelp JSON

5GB raw data

Python Processing

Split & transform

Amazon S3

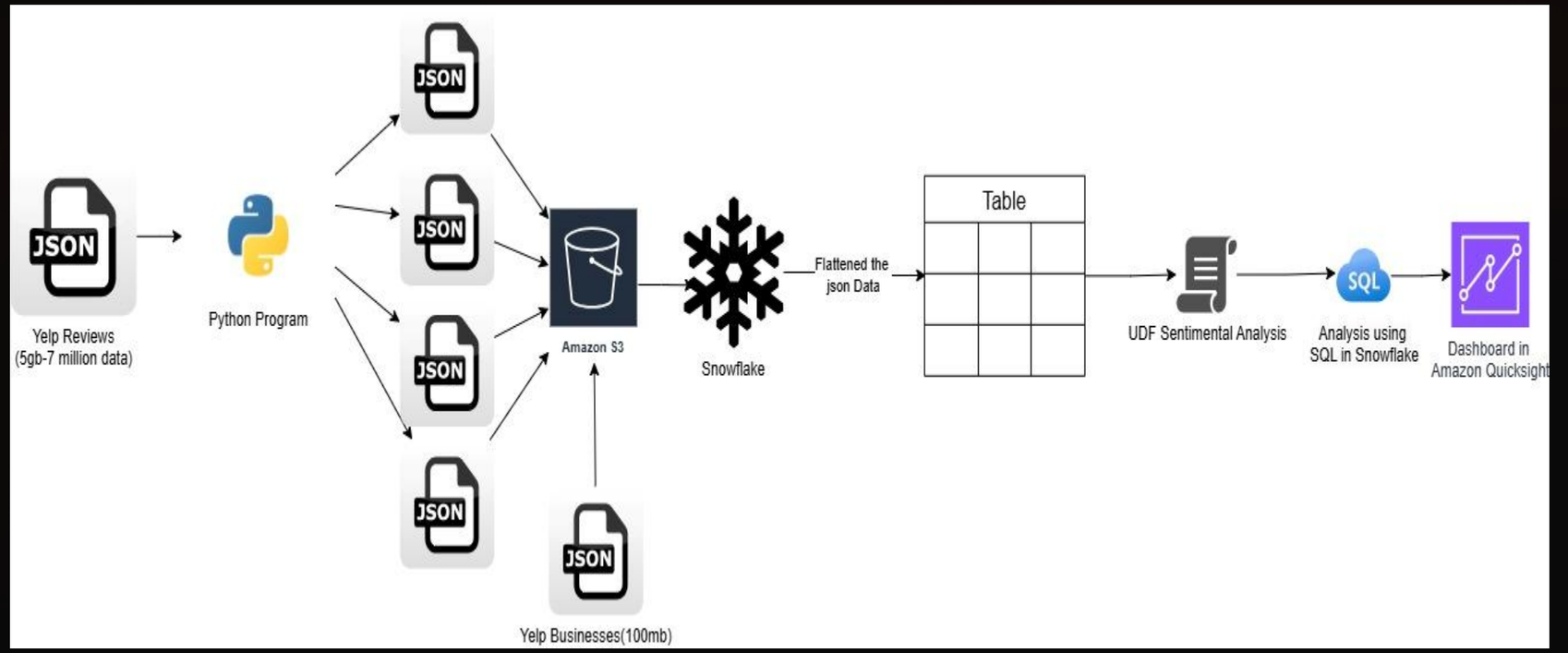
Cloud storage

Snowflake

Data warehouse

QuickSight

Visualization



Scalable pipeline designed to handle full dataset with UDF sentiment analysis and SQL-based insights generation.

Project Scope & Objectives

114,117 Businesses

Across 22 U.S. states and 1,416 cities

150,346 Reviews

Sampled from ~7M total reviews

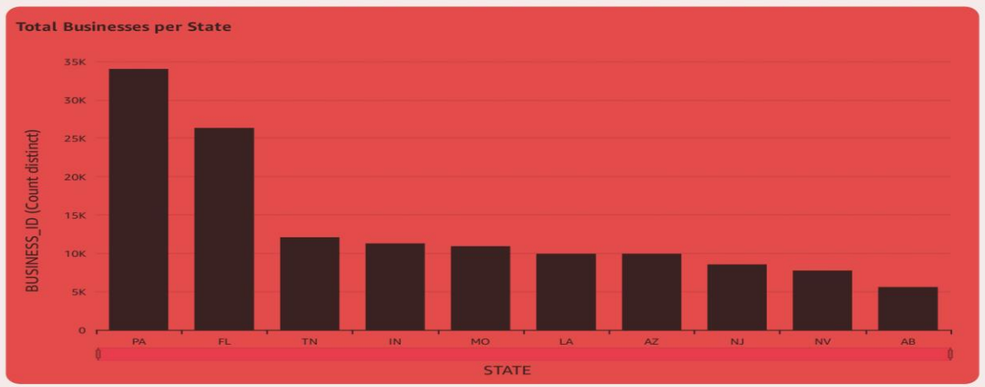
Key Insights

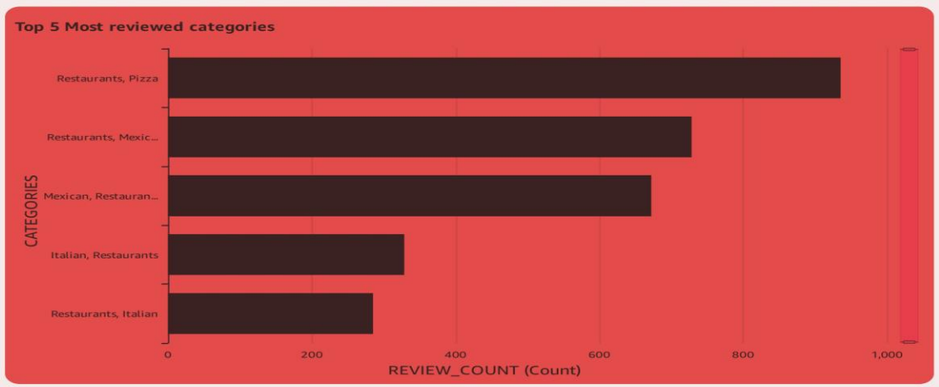
Business distribution, customer sentiment, and review trends

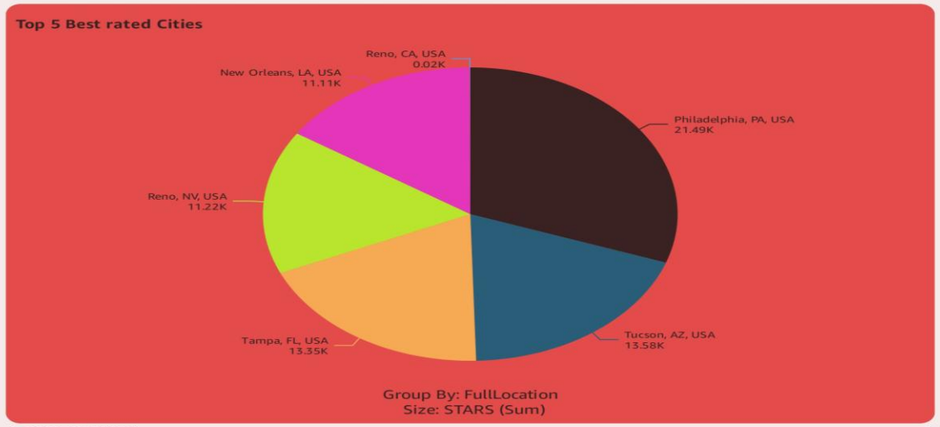
yelp Business

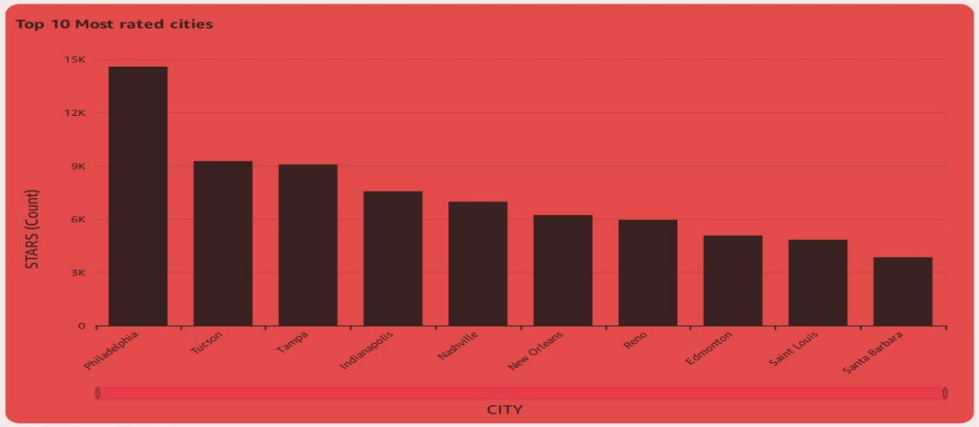
Yelp Businesses Analysis

Total Cities	Total Businesses	Total Categories	Total 5 Star Businesses	Total States
1,416	114,117	83,160	43,488	22

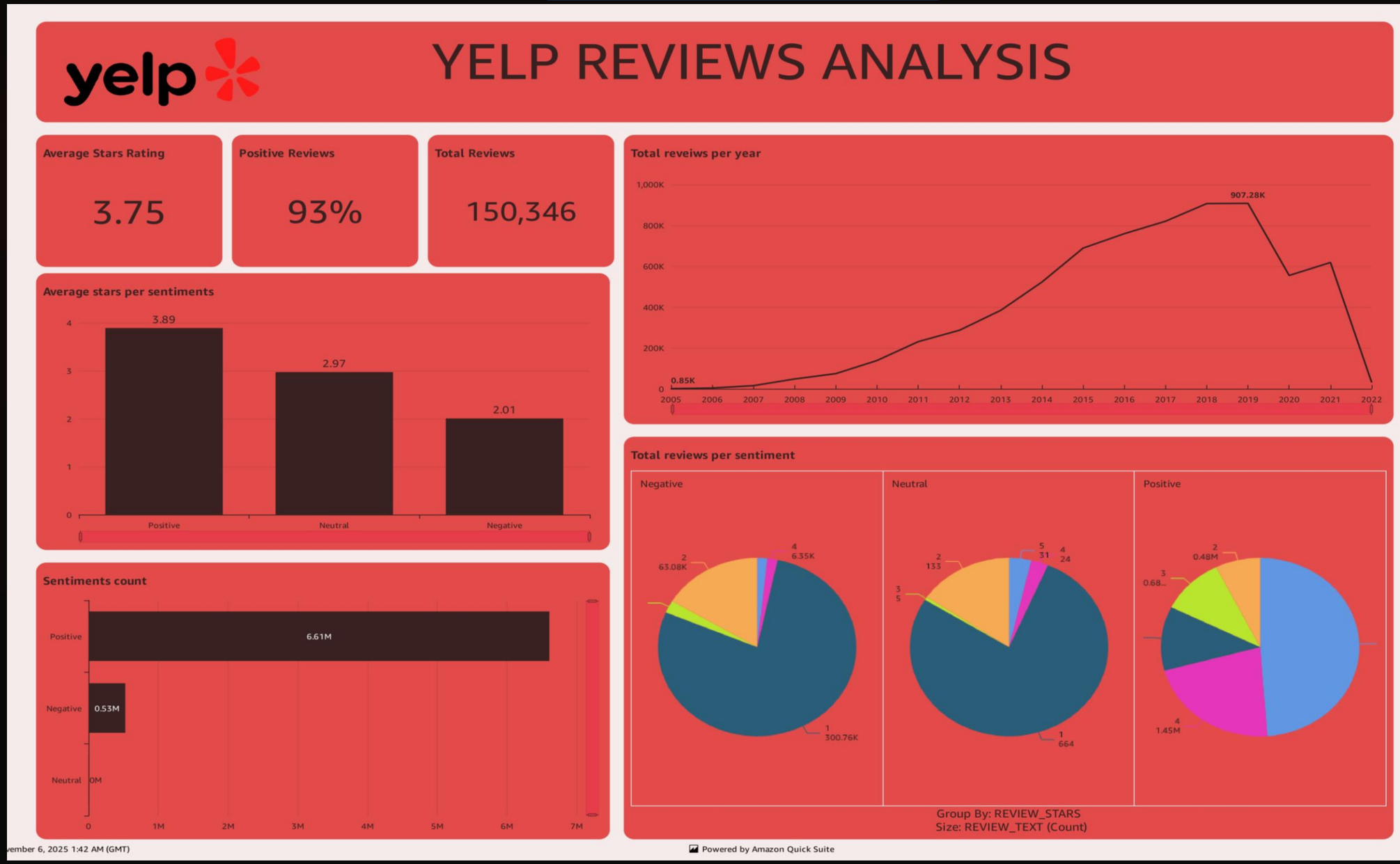








yelp Reviews



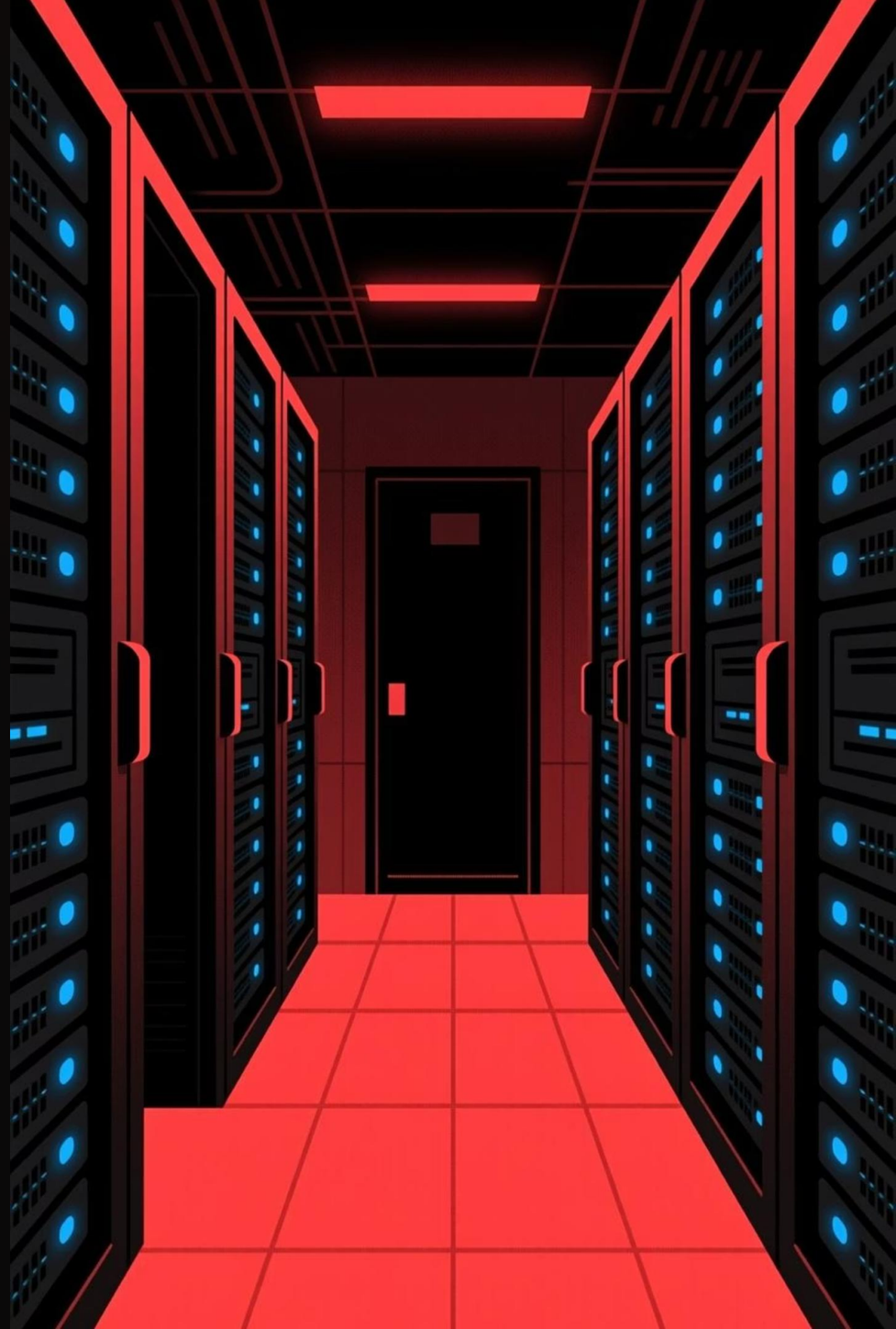
Dataset Overview

Business Data

- 100 MB file size
- 114,117 business records
- 83,160 unique categories
- 1,416 cities covered

Review Data

- 150,346 sampled reviews
- ~7M total reviews available
- Sentiment analysis applied
- Multi-year trend data



Business Distribution Insights

114K

Total Businesses

Across all states

1,416

Cities

Geographic coverage

22

States

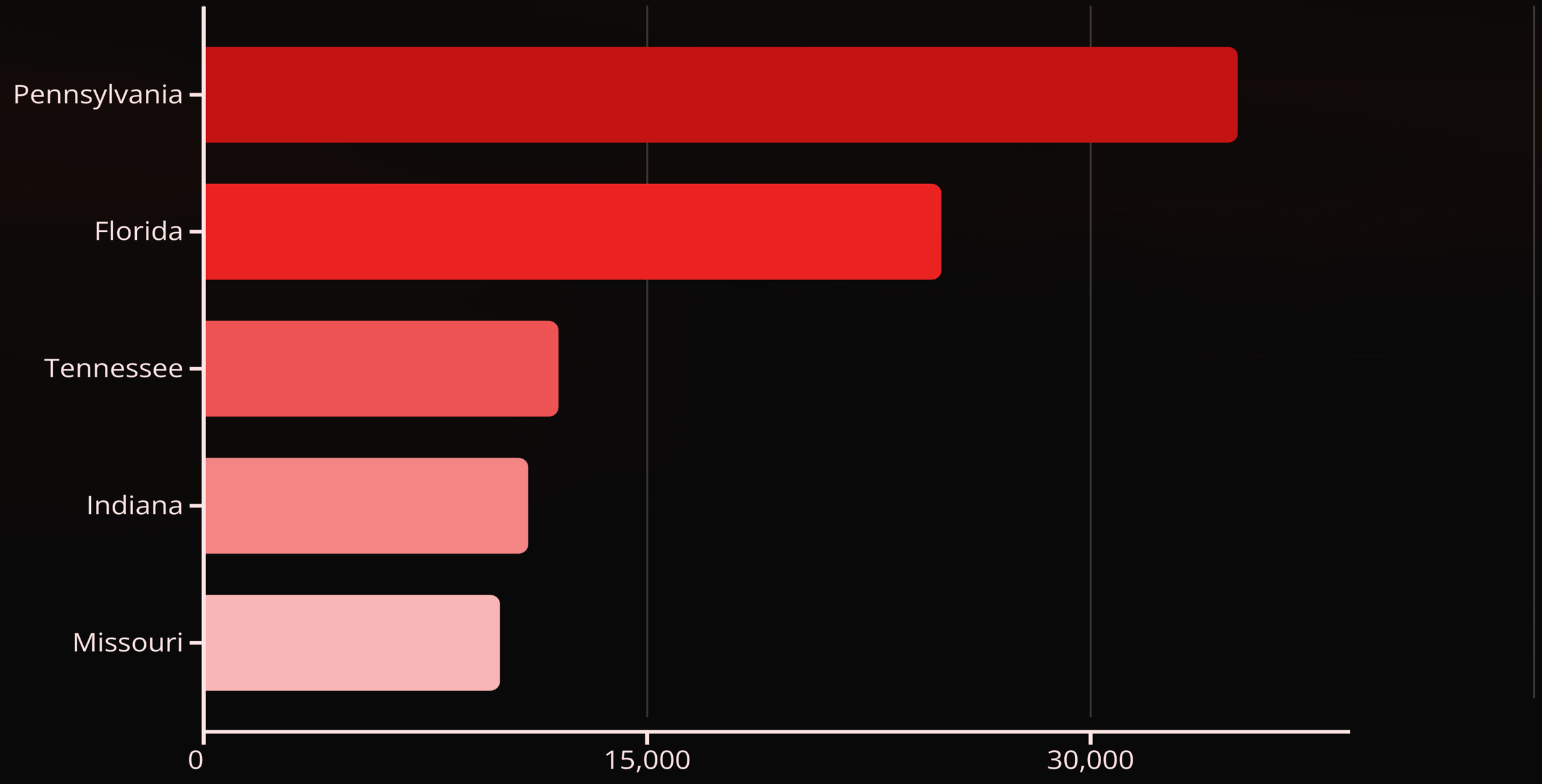
U.S. presence

43K

5-Star Businesses

Top-rated establishments

Top Performing States



Most Popular Business Categories

1

Restaurants, Pizza

Highest review volume across all categories

2

Restaurants, Mexican

Strong customer engagement and reviews

3

Mexican Restaurants

Popular dining choice nationwide

4

Italian Restaurants

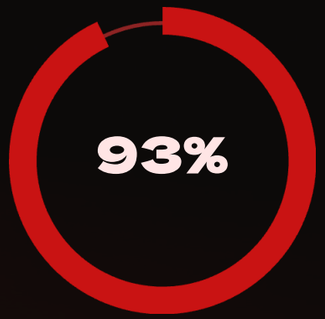
Consistent review activity

5

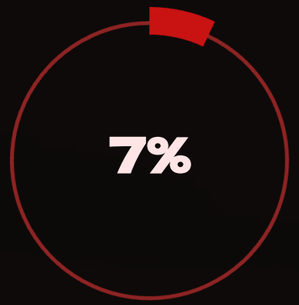
Restaurants, Italian

High customer satisfaction ratings

Sentiment Analysis Results



Positive Reviews



Negative Reviews

Neutral Reviews

Average Star Ratings by Sentiment

- Positive: 3.89 stars
- Neutral: 2.97 stars
- Negative: 2.01 stars

Overall Average: 3.75 stars

Yelp users demonstrate overwhelmingly positive sentiment, with 93% of reviews classified as positive.

Review Volume Trends Over Time

Key Takeaways & Conclusions

93% Positive Sentiment

Yelp users are highly satisfied, with positive reviews dominating the platform

Restaurants Dominate

Food service categories lead in both business count and review volume

PA Leads Geography

Pennsylvania has the highest business concentration with 35K establishments

Scalable Pipeline

Architecture ready to process full 7M dataset with 10 SQL queries powering insights

Made by Amazon QuickSight, Snowflake, AWS S3 & Python

THANK YOU