

# YELP Dataset

Sentiment Analysis

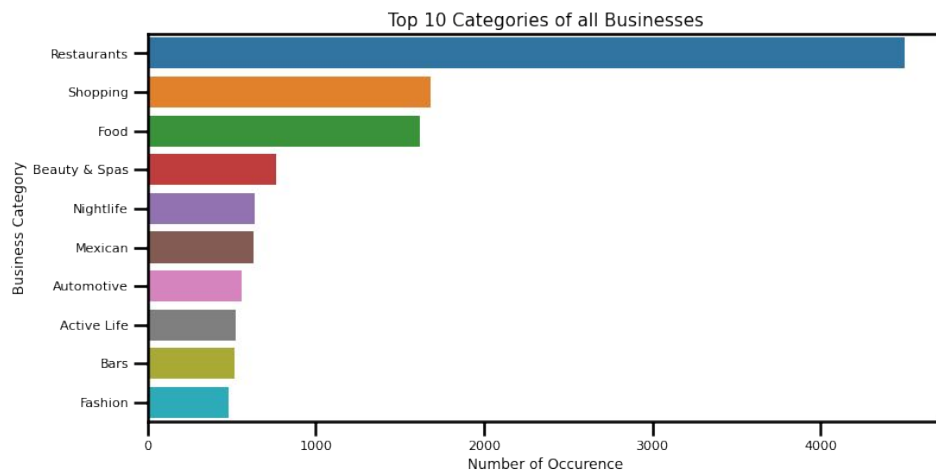
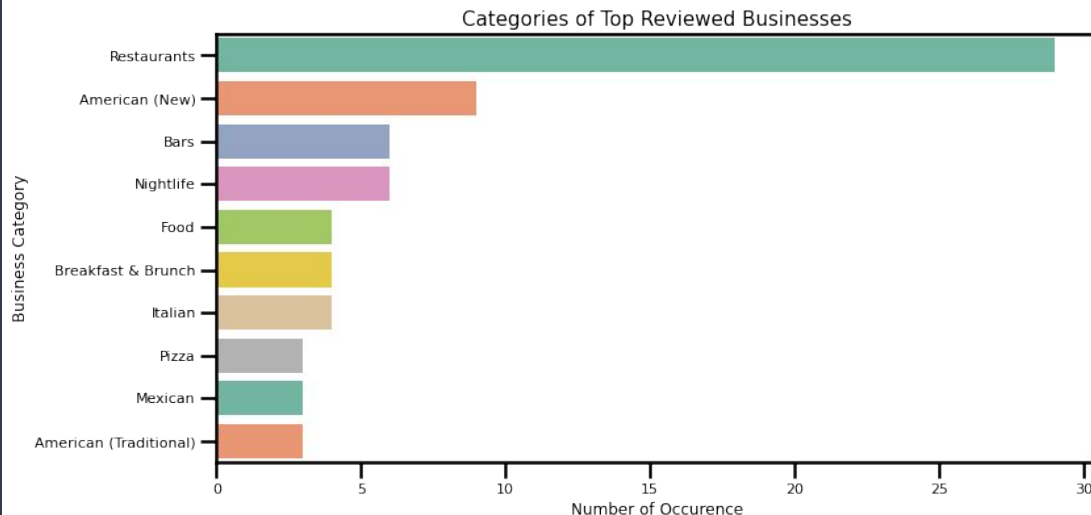
A large, dark blue, abstract shape that starts as a thin line on the left and expands into a wide, solid area at the bottom right, creating a modern, minimalist design element.

# GOAL OF THE PROJECT

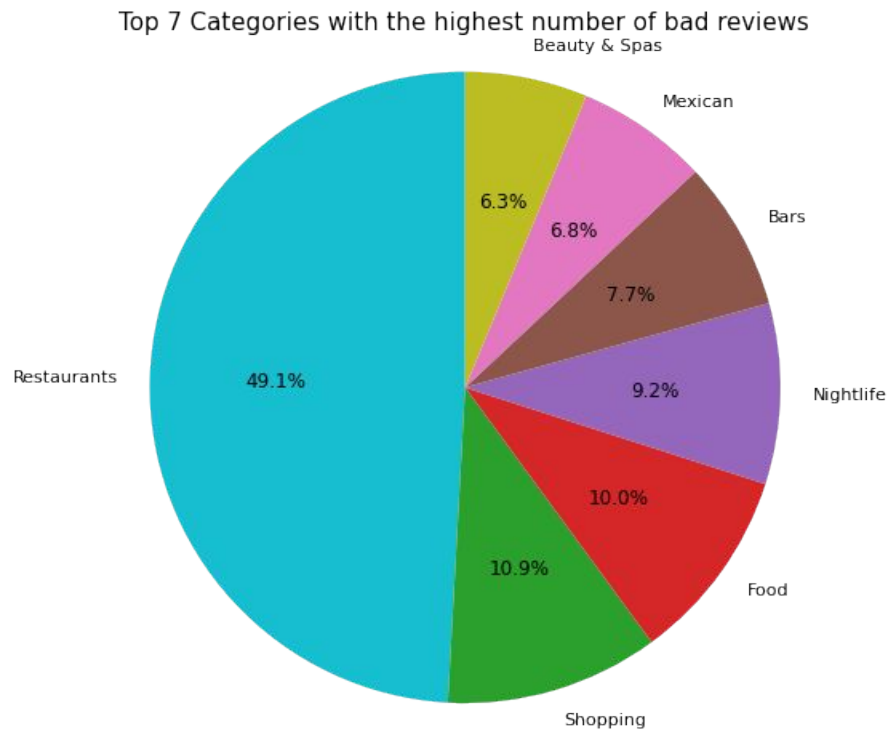
- Explore the dataset to gain valuable insights and answer the following:
  - What businesses are getting top reviews?
  - Which categories of businesses are getting top reviews?
  - How often do businesses get reviewed over time?
  - How do the categories of trending and top reviewed businesses differ?
  - Which business categories get bad reviews?
  - What are the most common words in bad reviews?
- Perform Sentiment Analysis
- Predict the Sentiment of a review

# Top categories that are being reviewed

It is clear that *Restaurants* are a dominant category when looking at top-rated businesses, as well as all categories in the dataset. However, looking at all categories *Restaurants* have a slightly smaller share, and is followed by *Shopping*, *Food*. There might be room for an assumption that people tend to review their experience with eating-out more than other consumer experiences. However, *Restaurants* and similar categories have a dominant share on Yelp, so this assumption might not hold entirely.



# Which categories get most bad reviews?



# Sentiment Analysis and Prediction

## NLTK Vader:

Accuracy: 0.71

	precision	recall	f1-score	support
negative	0.68	0.39	0.49	38245
neutral	0.23	0.10	0.14	35268
positive	0.76	0.93	0.84	155617
accuracy			0.71	229130
macro avg	0.55	0.47	0.49	229130
weighted avg	0.66	0.71	0.67	229130

## Various classifiers:

	Model Name	Train Accuracy	Validation Accuracy	Accuracy Difference
0	Logistic Regression	0.801329	0.788683	0.012646
0	Multinomial Naive Bayes	0.696545	0.687671	0.008874
0	Random Forest	0.760727	0.717962	0.042766
0	Decision Tree	0.711174	0.708625	0.002549
0	KNN	0.680715	0.680636	0.000080
0	AdaBoost	0.690552	0.690515	0.000036
0	XGBoost	0.825757	0.775546	0.050211

Thank you!

Contact me —>

**Website:** [www.adzictanja.com](http://www.adzictanja.com)

**GitHub:**

<https://github.com/adzict>

**LinkedIn:**

<https://www.linkedin.com/in/tanja-ad%C5%BEi%C4%87/>