



Comparing Kansas City and St. Louis Metropolitan areas

MATTHEW PERKINS

Why compare KC to STL

- ▶ Knowing similarities and differences helps decide which cities to move into
- ▶ Businesses can decide optimal location in each metro to open a location
- ▶ These two metro areas are often compared together

Goal

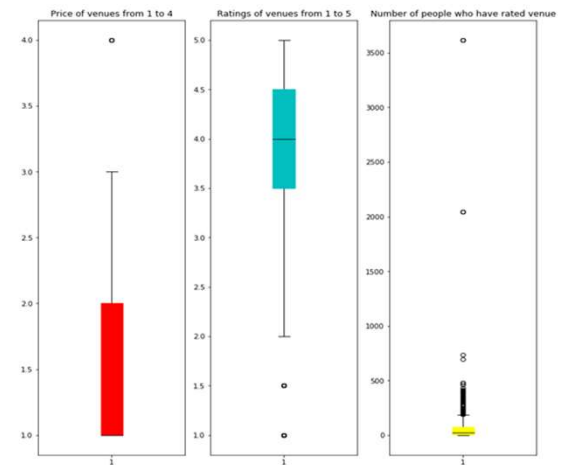
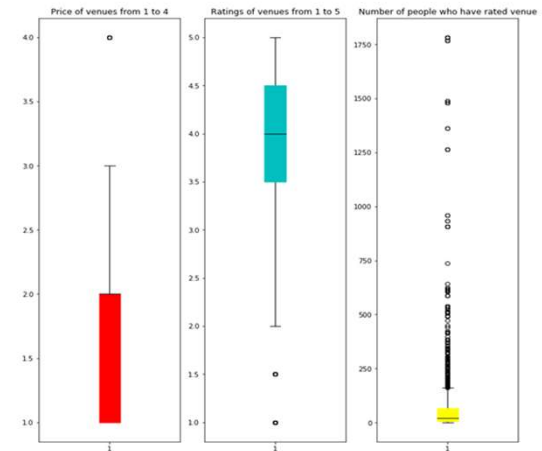
- ▶ Find groups of cities in each metro area that are alike
 - ▶ Price, quality, and type of venues are the factors

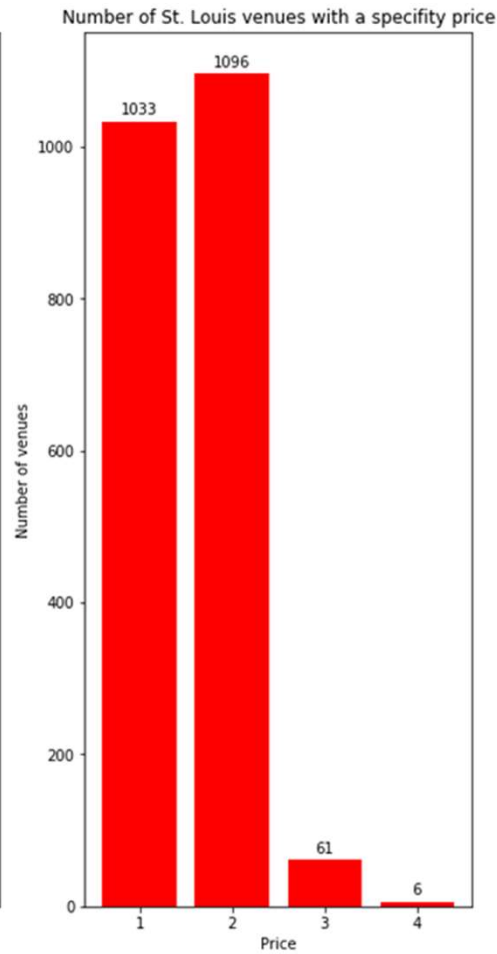
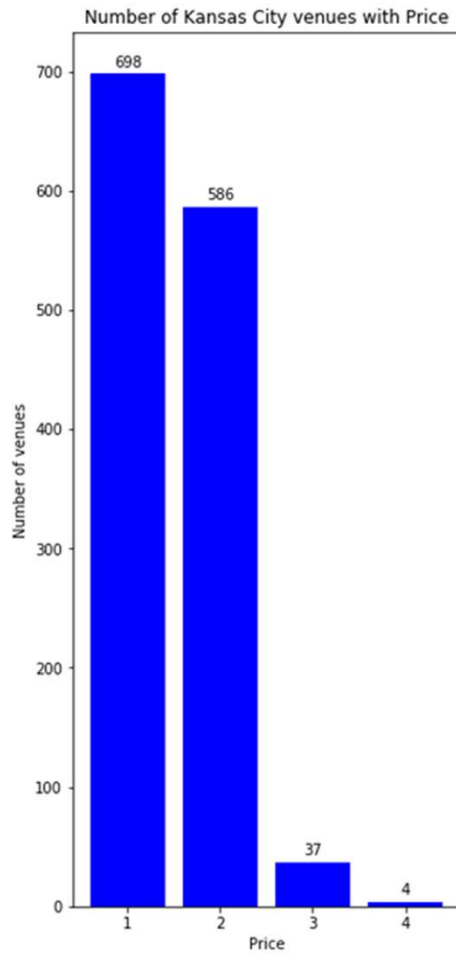
Data Collection and Cleaning

- ▶ List of cities in each metro area comes from their Wikipedia pages
- ▶ Venue data collected from Foursquare and Yelp APIs
 - ▶ Gathered price, rating, and number of people who've rated the venue from Yelp
- ▶ Removed venues with no rating
- ▶ Replaced empty price of venues with most frequently occurring price in the venue's city

Visualizing Data

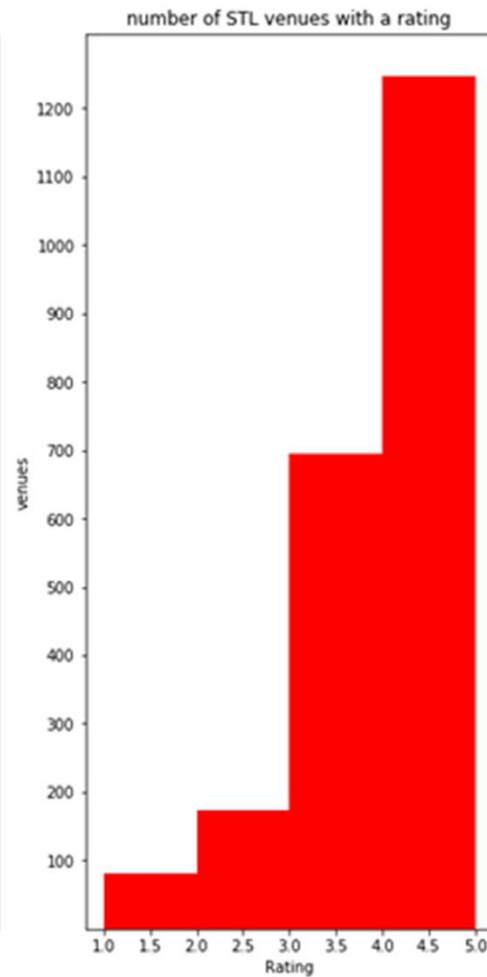
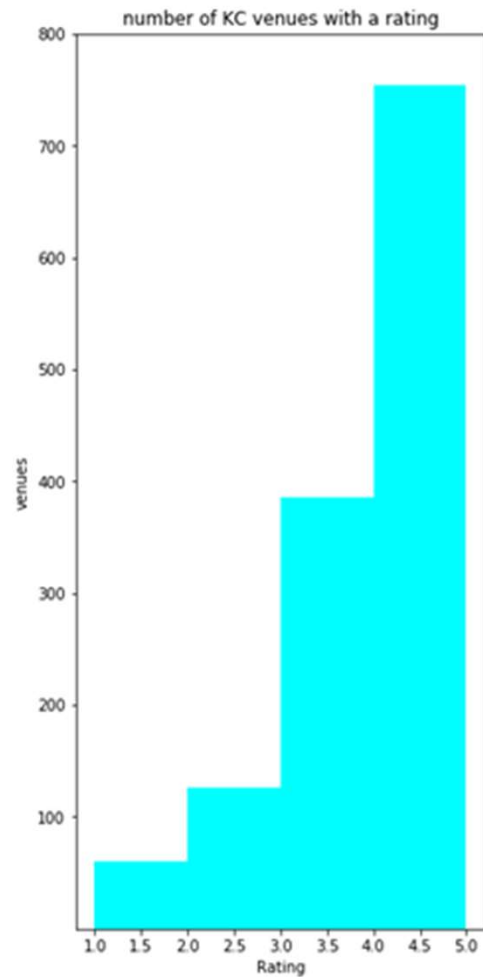
- ▶ Created a set of boxplots for price, rating, and number of people who've rated the venue for each metro area
 - ▶ St. Louis is subplot on top.
 - ▶ Kansas City on bottom
- ▶ Price and rating have a few outliers
- ▶ Price is right skewed
- ▶ Rating is left skewed.





Visualizing Data

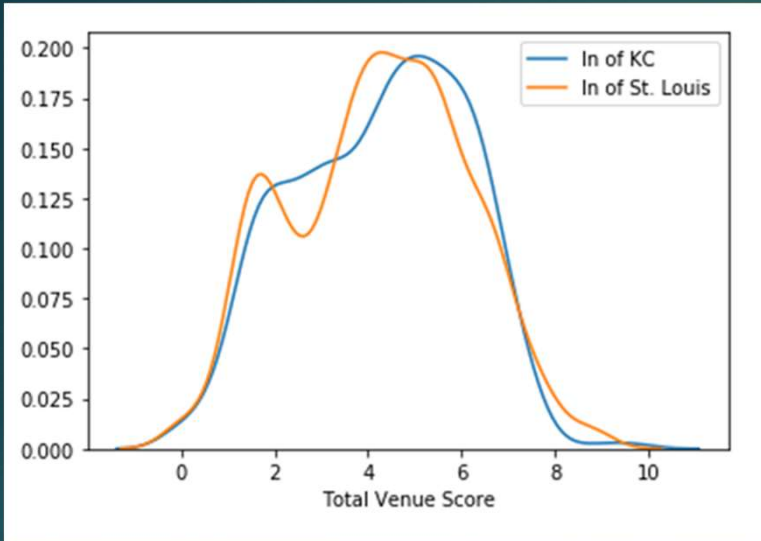
CREATED BAR GRAPHS FOR PRICE VARIABLE



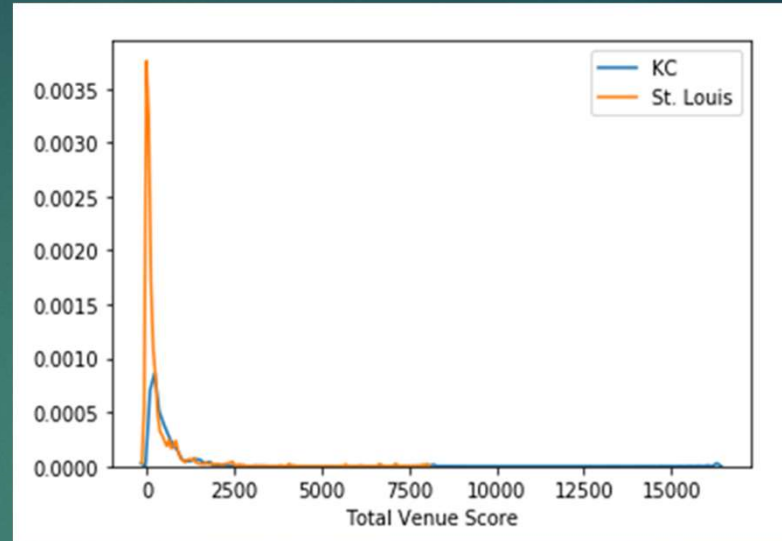
Visualizing Data

CREATED HISTOGRAMS FOR RATING VARIABLE

Natural log of total score



Total score



Feature Engineering

- ▶ Multiplied rating by number of people who rated the venue to get total score for that venue
- ▶ Performed natural log transformation on total score to make a more normal distribution

Grouping the cities

- ▶ Kmeans clustering algorithm used to cluster cities
- ▶ 10 clusters were used to group cities
 - ▶ One cluster (cluster 8 in code) contained one city (an outlier)
 - ▶ 5 clusters (clusters 0,1,7,8,9) contained cities belonging to outskirts of metro areas
 - ▶ 2 clusters (clusters 2 and 5) belonged to downtown areas
 - ▶ 3 clusters (clusters 3,4, and 6) belonged to the suburbs
- ▶ Price and total score were major factors in groupings

Conclusion

- ▶ Kansas City and St. Louis are very similar
- ▶ St. Louis slightly more expensive
- ▶ Similar quality between the two