

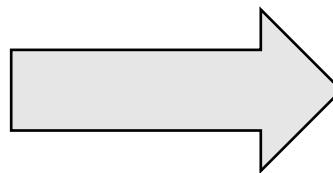
Analyzing destinations by gas price, temperature, or venue type frequency

Andrew Samarakone
October 22nd 2018

Why?

- Try to figure out alternative destinations for a vacation-seeking customer

Happy vacation-seeking customer



Successful business



Sources of Data

- Bureau of Transportation Statistics (BTS)

Bureau of Transportation Statistics

- Wikipedia



- Gas Buddy

- Weather Underground



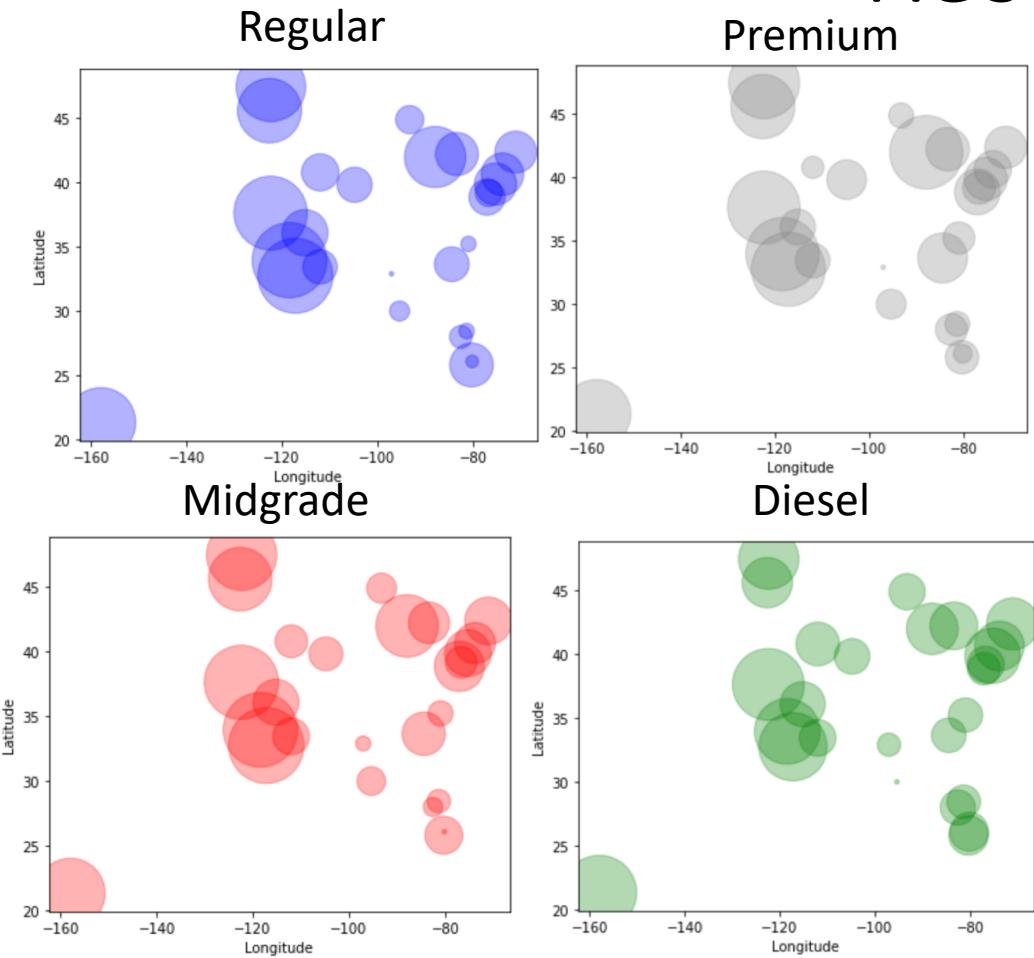
- Foursquare



Main Methodology

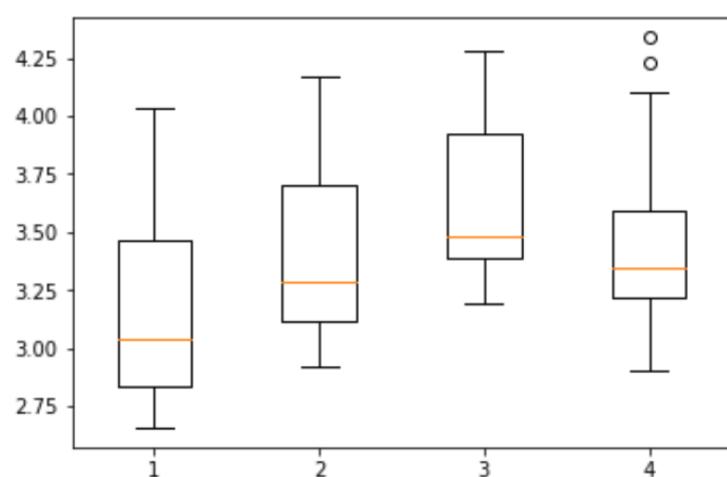
- Plots
 - 1. Box
 - 2. Scatter / Bubble
 - 3. Folium map
- Machine Learning Clustering Algorithms
 - 1. K-Means
 - 2. DBSCAN

Results/Discussion

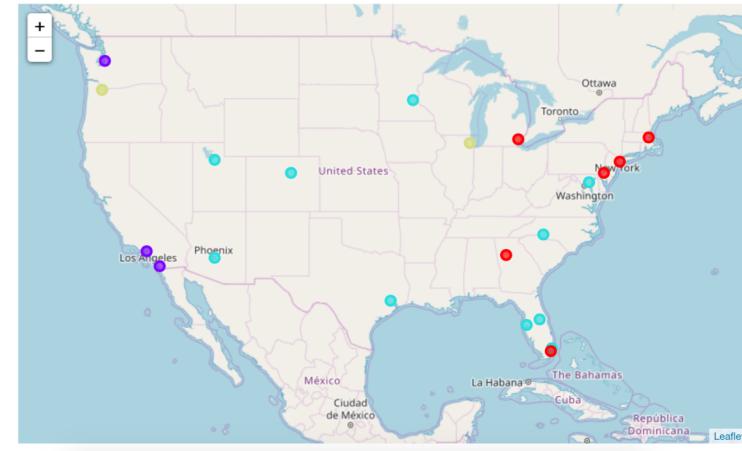


Phoenix, and Denver seem to have similar gas prices
 Boston and Philadelphia seem to have similar gas prices
 Baltimore, Atlanta, Tampa, Minneapolis, and Orlando
 seem to have similar gas prices

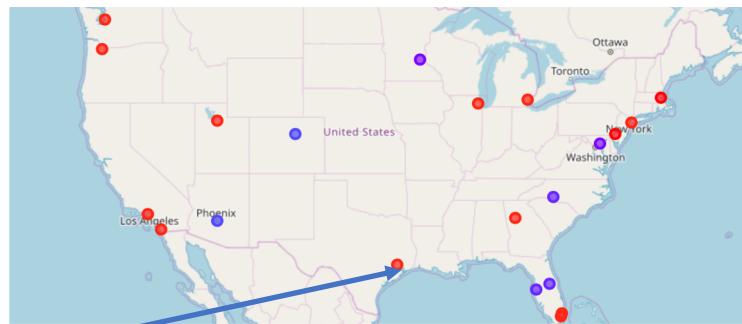
Example of a recommendation:
 If someone wants to visit a place with
 similar gas prices to Minneapolis,
 suggest Orlando



	Regular Gas Price Estimate [\$]	Premium Gas Price Estimate [\$]	Diesel Gas Price Estimate [\$]
count	26.000000	26.000000	26.000000
mean	3.169135	3.408512	3.473265
std	0.455359	0.400248	0.369642
min	2.655714	2.916667	2.905000
25%	2.833571	3.115833	3.220714
50%	3.034786	3.286875	3.347875
75%	3.468500	3.700562	3.596500
max	4.032857	4.175000	4.340000

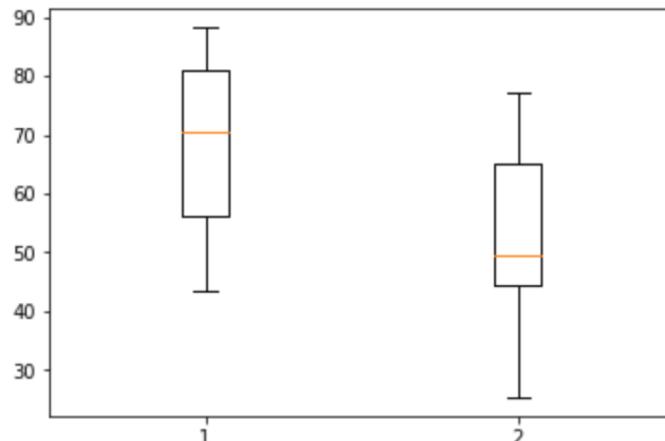


DBSCAN (eps = 0.29, min_samples = 2)



Most of the red dots are outliers
 Exceptions: Philadelphia & Boston

Results/Discussion (cont.)



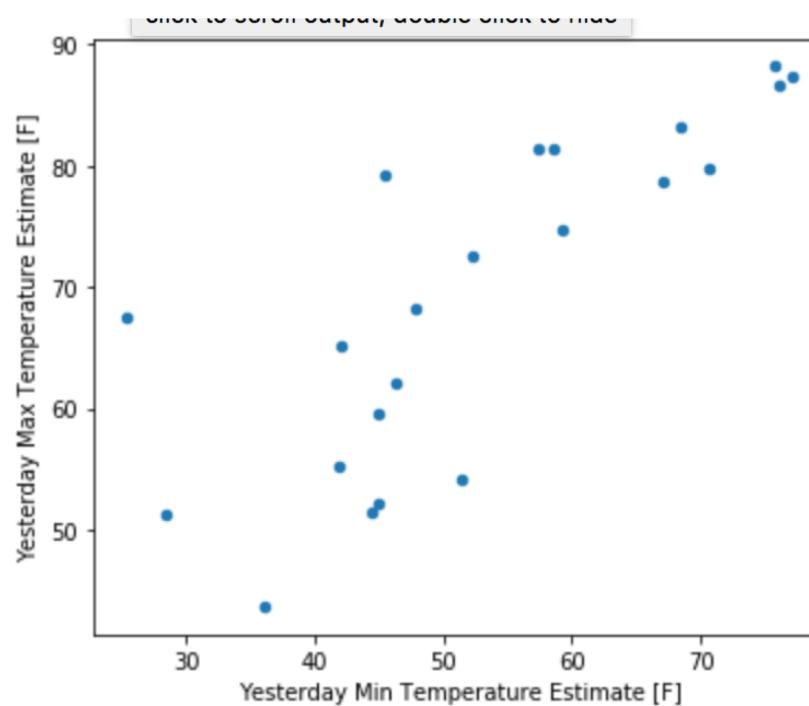
Yesterday Max Temperature Estimate [F]

	Yesterday Max Temperature Estimate [F]
count	22.000000
mean	69.254545
std	13.810649
min	43.600000
25%	56.275000
50%	70.400000
75%	81.050000
max	88.200000

Yesterday Min Temperature Estimate [F]

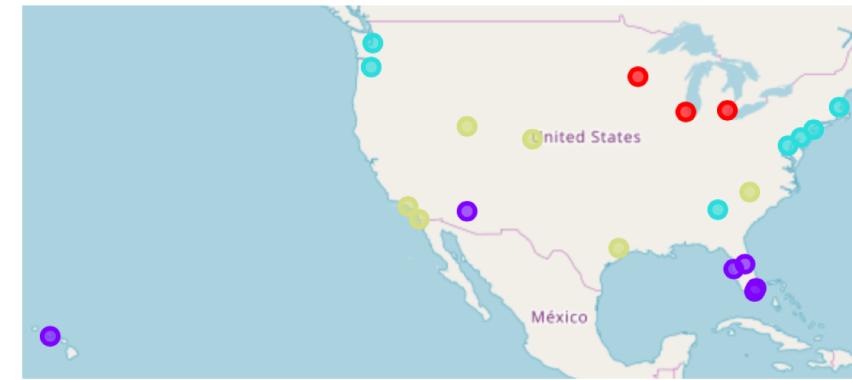
	Yesterday Min Temperature Estimate [F]
count	22.000000
mean	52.831818
std	14.967374
min	25.400000
25%	44.550000
50%	49.650000
75%	65.150000
max	77.200000

Min/Max Temperature Estimates for Yesterday

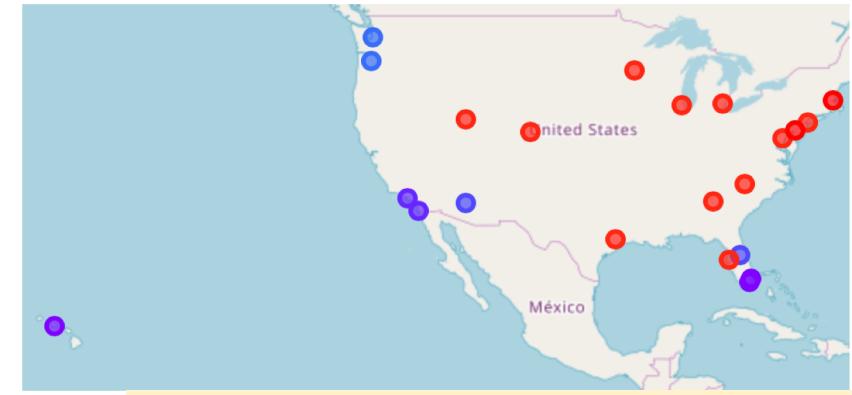


Check to see output, double click to hide

K Means (k = 4)



DBSCAN (eps = 0.3, min_samples = 2)



Seattle and Portland seem to have similar temperature ranges yesterday
Boston and Philadelphia seem to have similar temperature ranges yesterday
San Diego and Los Angeles seem to have similar temperature ranges yesterday
Honolulu, Tampa, Miami, and Fort Lauderdale seem to have similar temperatures yesterday



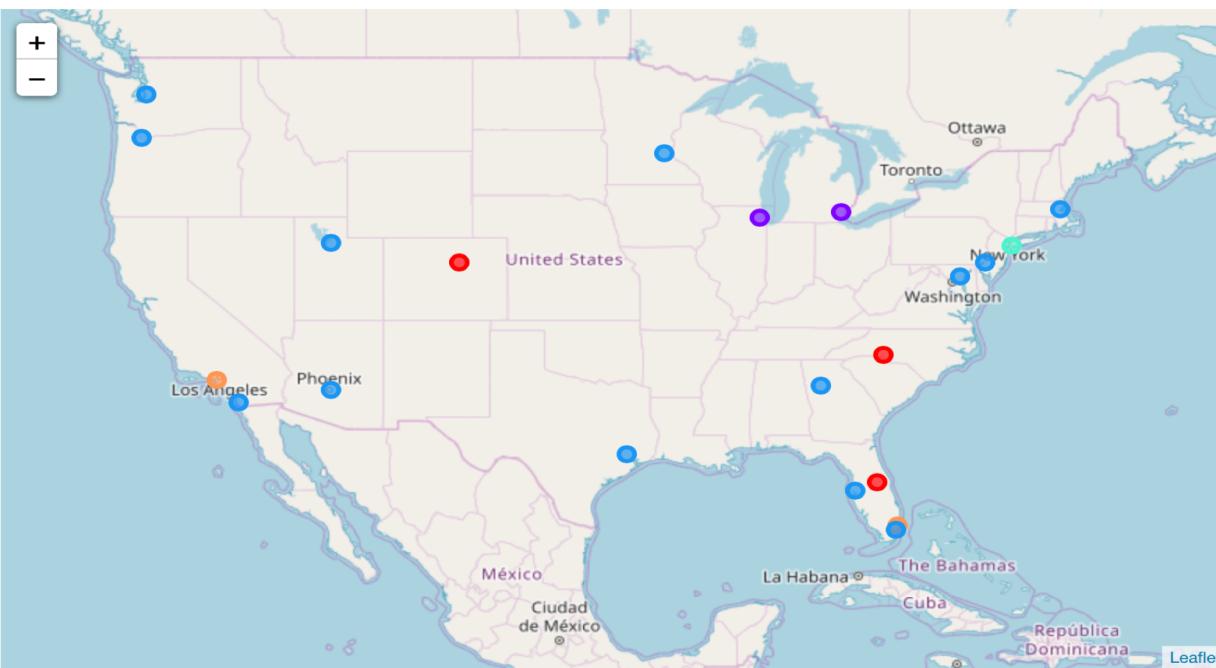
Example of a recommendation:

If someone trying to book a flight for immediate departure wants to visit a place with temperature ranges like Miami, suggest Honolulu

Results/Discussion (cont.)

Venue Type Frequency

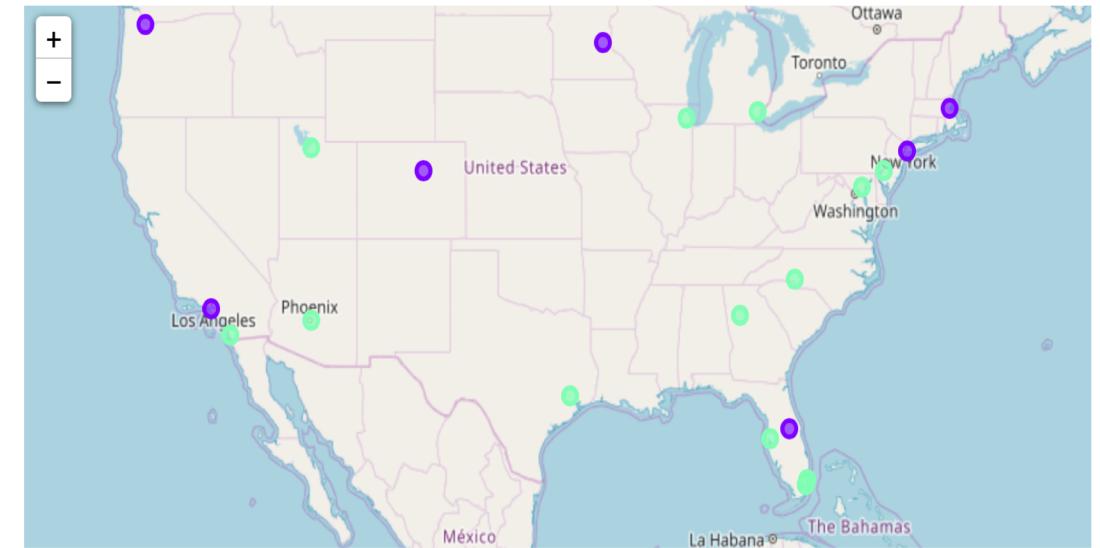
K Means ($k = 6$)



Vicinity of each large airport hub from Phoenix, Los Angeles, Salt Lake City, Houston, Atlanta, Tampa, Miami, Baltimore, and Philadelphia seem to have similar venue type frequencies

Outliers are purple

DBSCAN ($\text{eps} = 21, \text{min_samples} = 4$)



Example of a recommendation:

If someone wants to visit Miami just because they liked the type of venues around, consider suggesting Phoenix (assuming they don't care about other factors)

Conclusion

- **Need to improve project**
 - Try to take into account other variables such as humidity, wind speed, UV, recorded incidents of flu/cold, bus pass costs, traffic, venue frequency of big brand names (e.g. Starbucks), frequency of slang words in Venue tips, etc.
 - Maybe look at historic data (e.g. historic temperature data) and make predictions using ML (e.g. Logistic Regression). For example, predict the probability that the temperature for the day the customer plans to arrive is between X and Y degrees Fahrenheit.