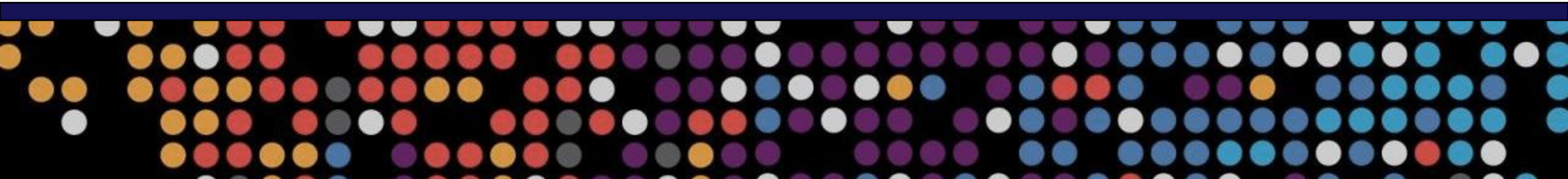
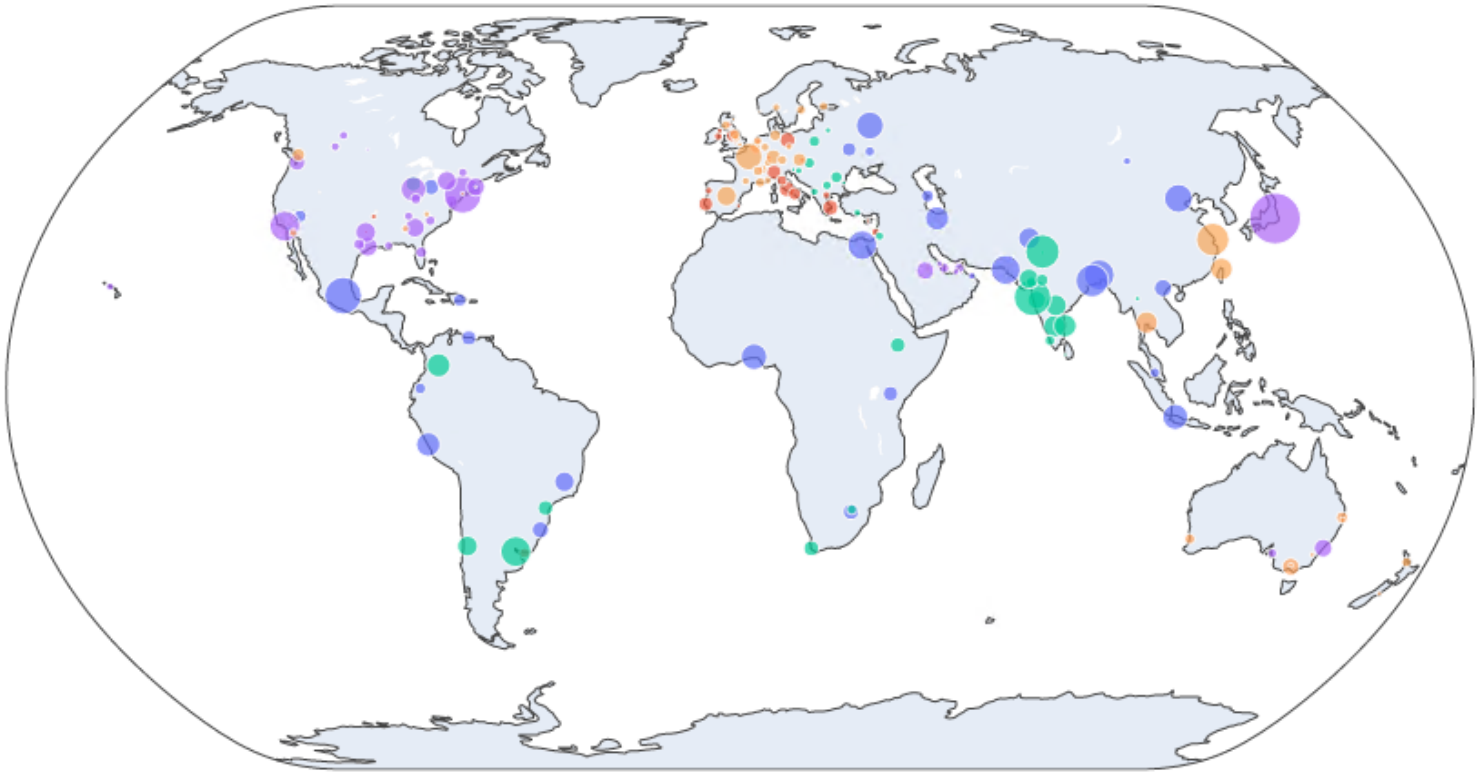
Participant: Keyla Gonzalez

City Search Tool Using Machine Learning Algorithms



Introduction

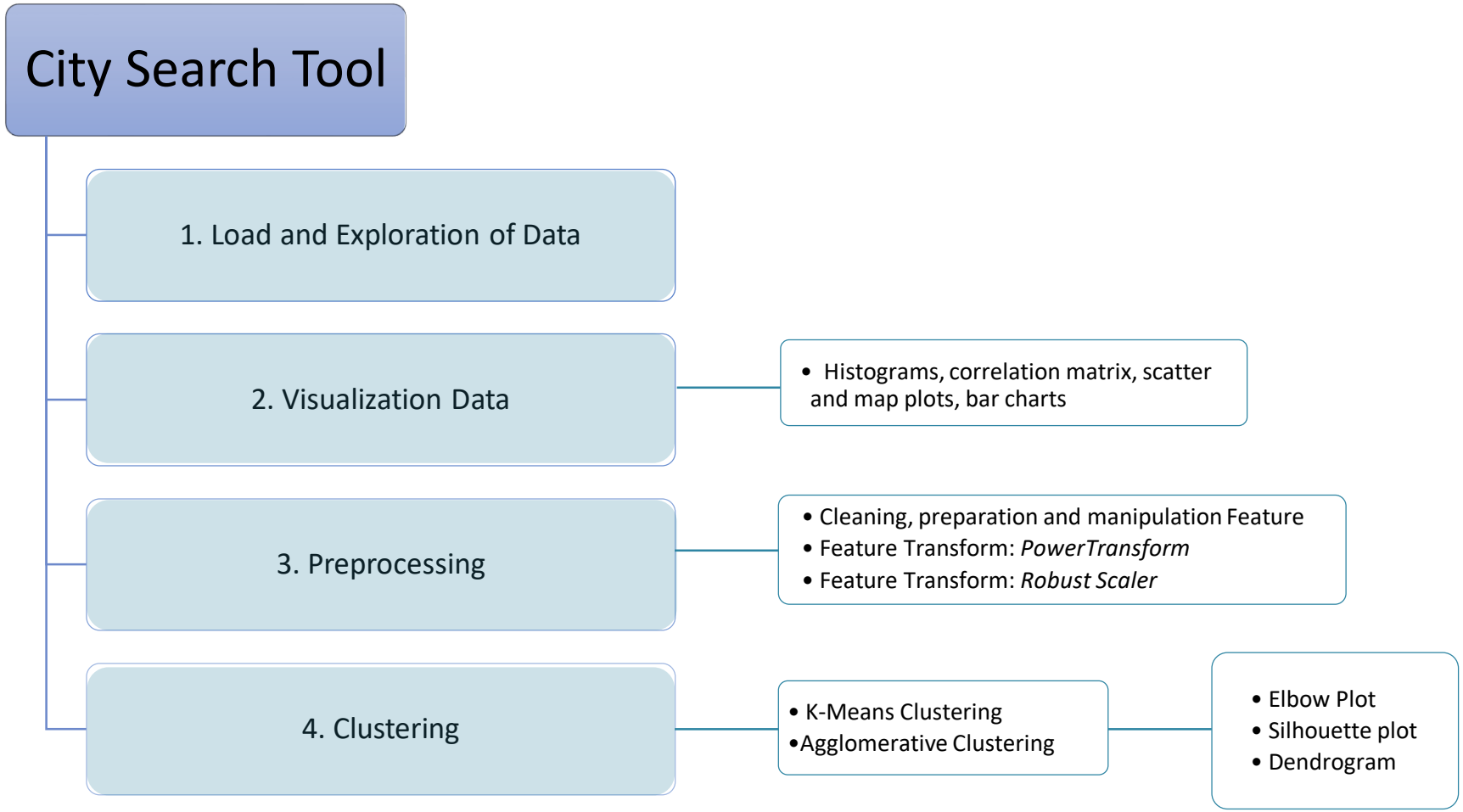


Agglomerative clustering

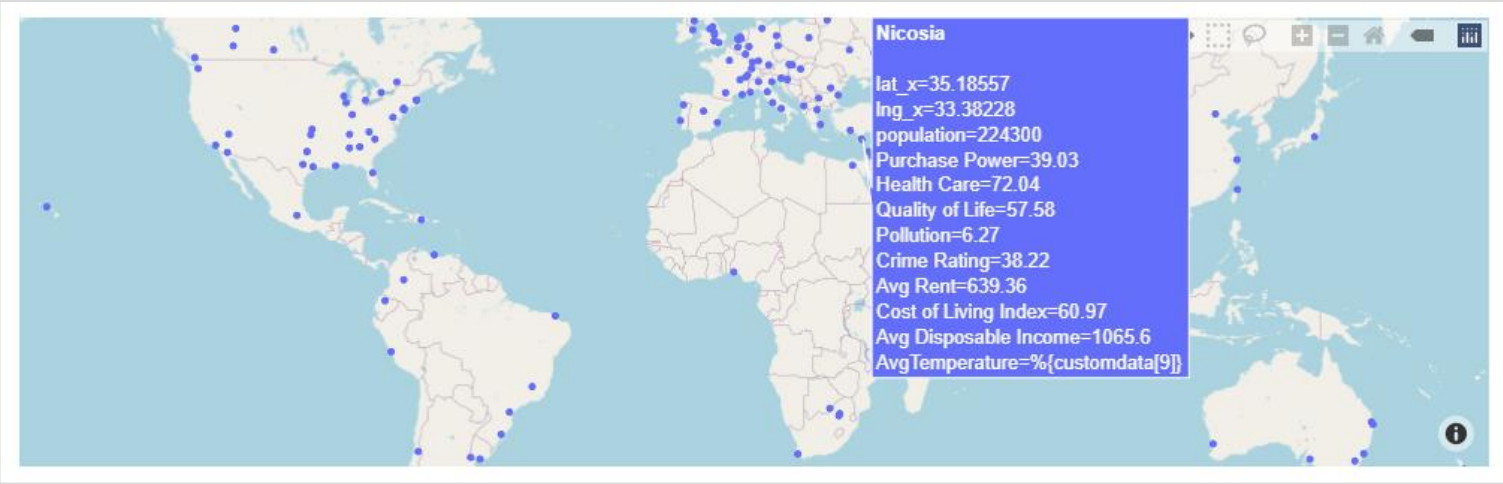
Moving to a foreign country is probably one of the biggest transitional steps in life

- Stressful
- Too many choices
- Friends and family
- Income/expenses
- Cultural shock

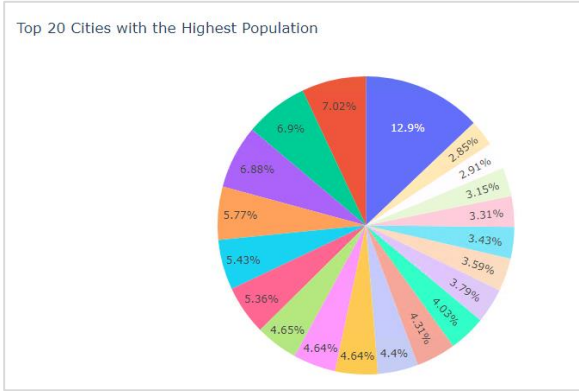
Workflow


City Search Tool
1. Load and Exploration of Data
2. Visualization Data
• Histograms, correlation matrix, scatter and map plots, bar charts
3. Preprocessing
• Cleaning, preparation and manipulation Feature
• Feature Transform: *PowerTransform*
• Feature Transform: *Robust Scaler*
4. Clustering
• K-Means Clustering
•Agglomerative Clustering
• Elbow Plot
• Silhouette plot
• Dendrogram

Results - Dataset Description



City	Movehub Rating	Purchase Power	Health Care	Pollution	Quality of Life	Crime Rating	lat_x	lng_x	Gasoline	Avg Rent	Avg Disposable Income	population	Rank	Cost of Living Index	Cost of Living Plus Rent Index
Aberdeen	81.89	49.70	82.86	34.31	76.77	24.22	57.149717	-2.094278	1.37	1195.74	1743.78	189364.0	107.0	73.02	51.69
Abu Dhabi	86.40	68.03	48.02	53.42	80.80	10.86	24.453884	54.377344	0.30	1779.93	2135.92	603492.0	269.0	61.94	60.28
Addis Ababa	59.88	6.38	63.89	85.59	28.41	26.04	8.980603	38.757760	0.72	653.77	124.22	3100000.0	330.0	49.96	39.93
Adelaide	87.29	72.03	56.25	12.01	91.54	41.32	-34.928499	138.600746	0.95	1382.26	2911.69	1145000.0	110.0	72.92	53.54
Ahmedabad	76.16	33.69	61.67	68.21	57.01	18.18	23.022505	72.571362	0.85	193.08	301.69	5375000.0	519.0	24.24	15.22
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
Victoria	81.44	52.91	79.63	68.93	80.87	42.45	-37.471308	144.785153	0.83	1083.77	1657.53	289625.0	209.0	66.17	52.88
Vienna	81.84	51.21	79.86	22.39	77.21	27.45	48.208174	16.373819	1.25	1248.90	1619.72	2400000.0	135.0	70.86	52.70
Vilnius	72.45	24.17	77.31	82.08	64.19	27.93	54.687156	25.279651	1.19	493.78	457.29	542366.0	340.0	48.31	33.51
Warsaw	76.76	35.77	63.33	86.16	51.82	32.03	52.229676	21.012229	1.18	726.59	664.31	1707000.0	376.0	42.32	32.04
Zagreb	73.84	26.71	52.78	83.45	47.59	32.29	45.815011	15.981919	1.23	504.36	605.23	722526.0	317.0	52.20	34.44



Dataset Information

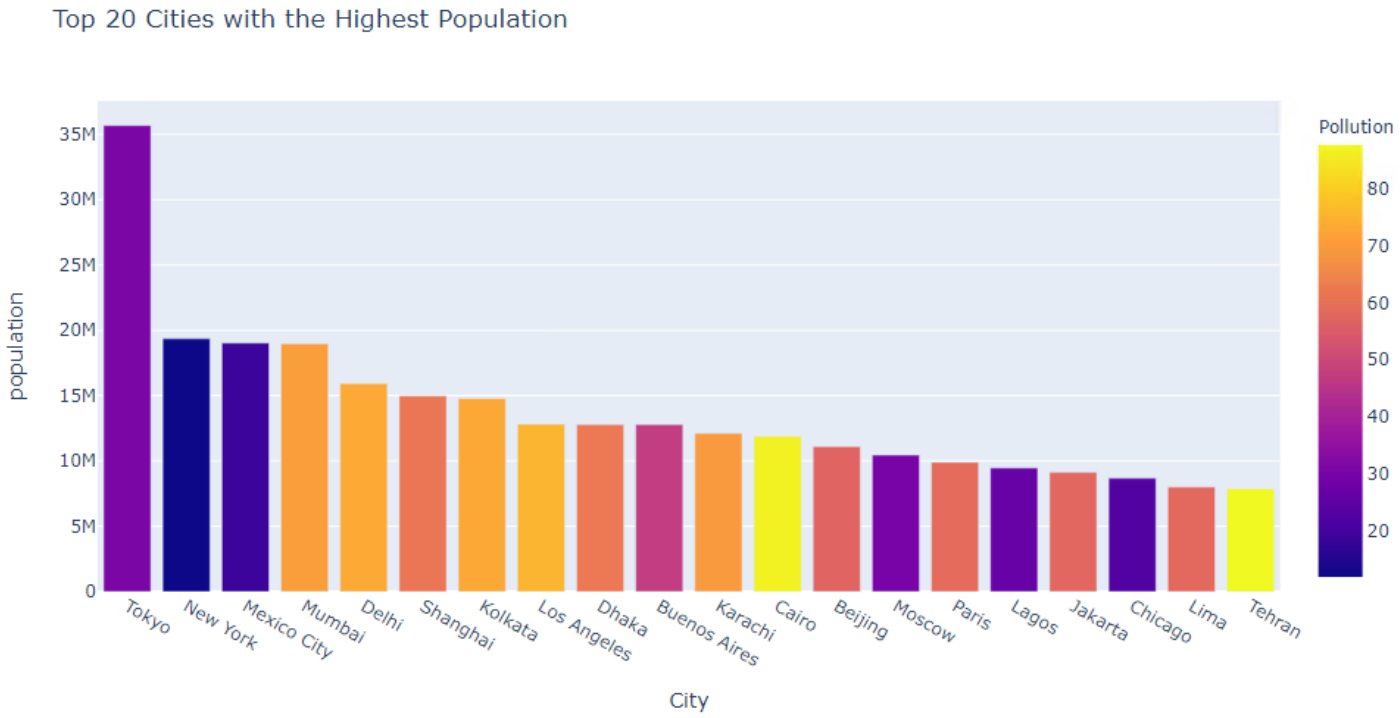
- 5 data sets
- Features:
  - Cities / Countries
  - Health Care
  - Pollution
  - Crime Rating
  - Gasoline
  - Population
  - Groceries Index
  - Average Temperature

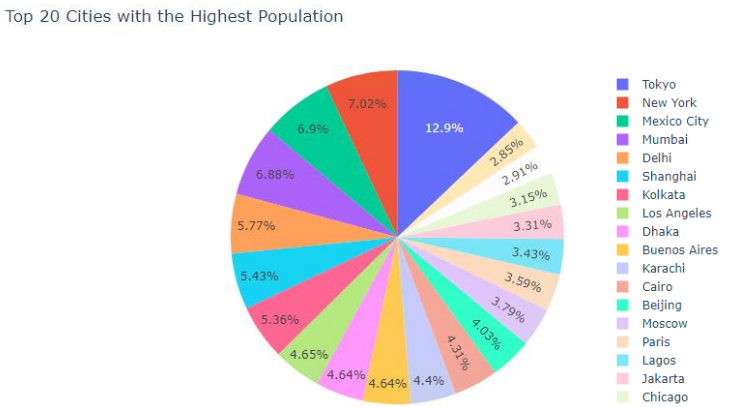
Goal → Suitable city to move (priorities/personality traits)

TAMU Datathon

Results - Visualization and Exploration

✓ **Most populated cities**



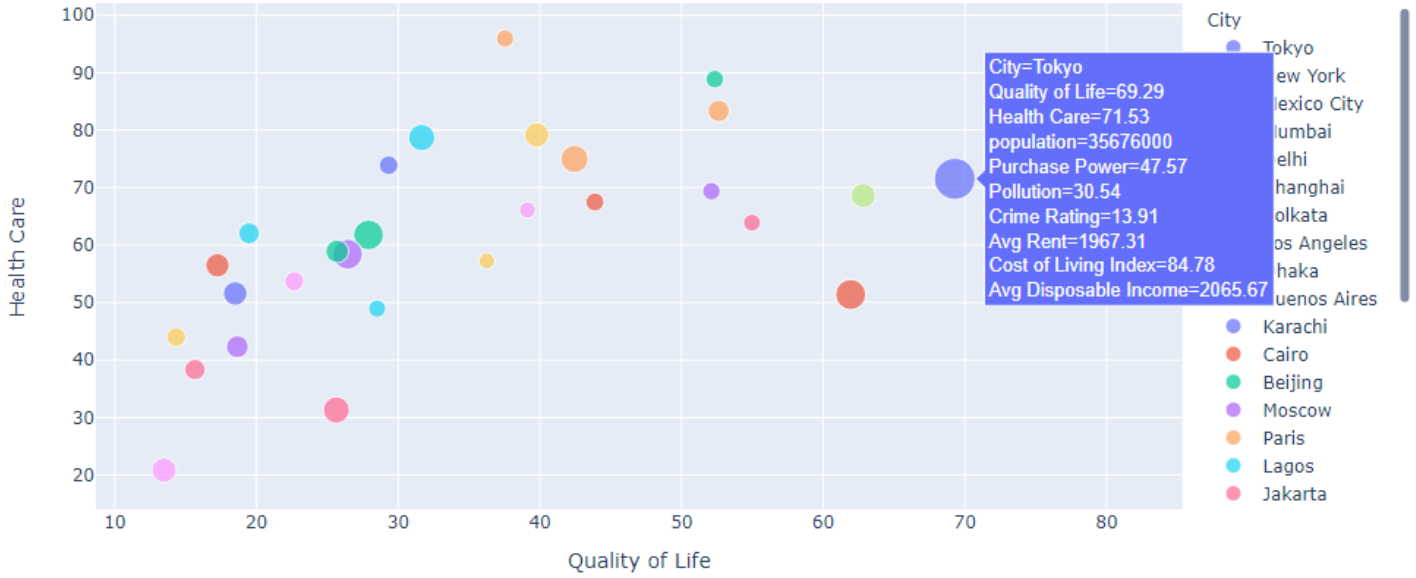


TAMU Datathon

Results - Visualization and Exploration

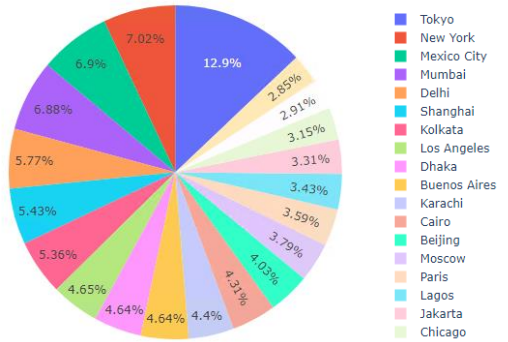
✓ **Most populated cities**

Top 30 Cities with the Highest Population



Health care vs Quality of life

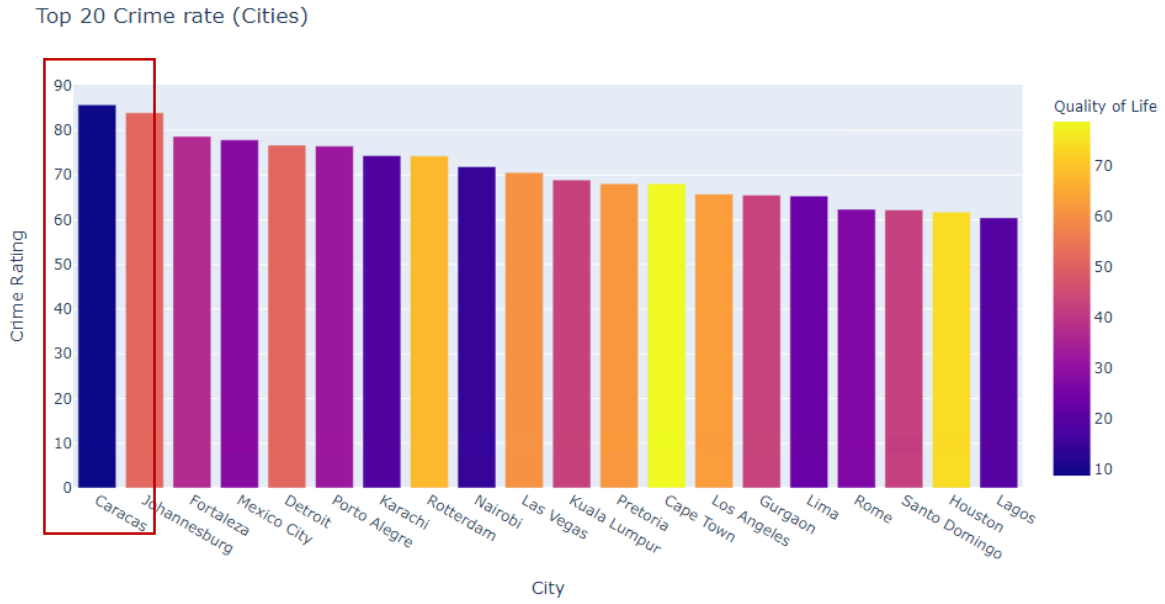
Top 20 Cities with the Highest Population



TAMU Datathon

Results - Visualization and Exploration

✓ **Top crime rate**

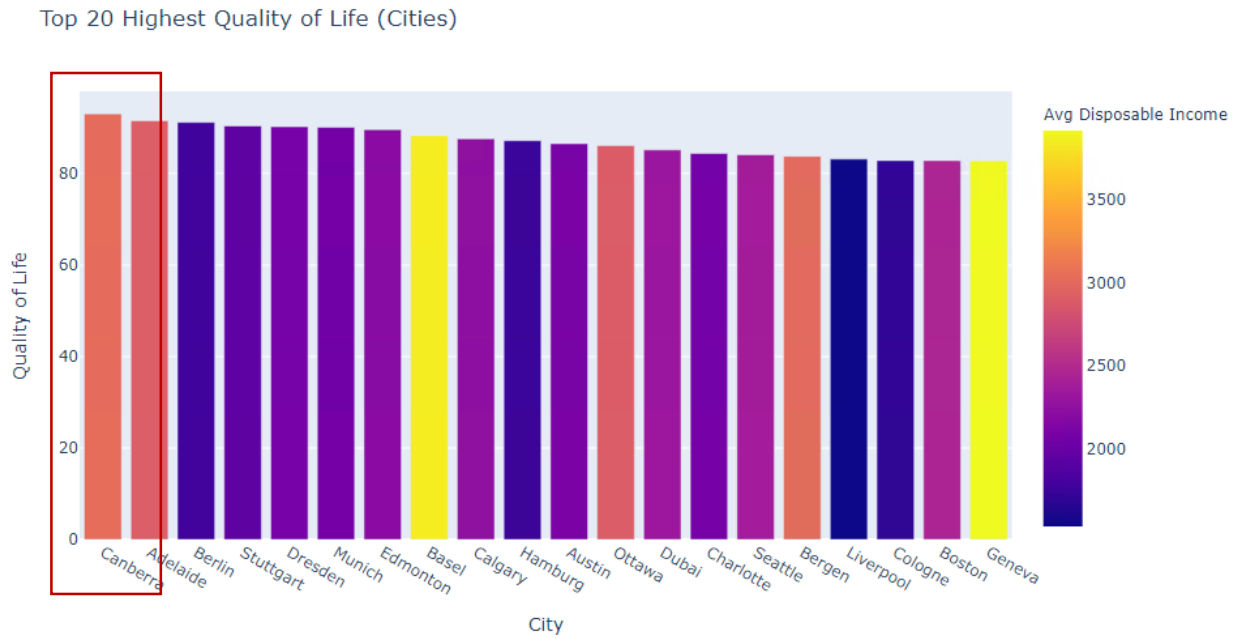


Crime rate vs Quality of life

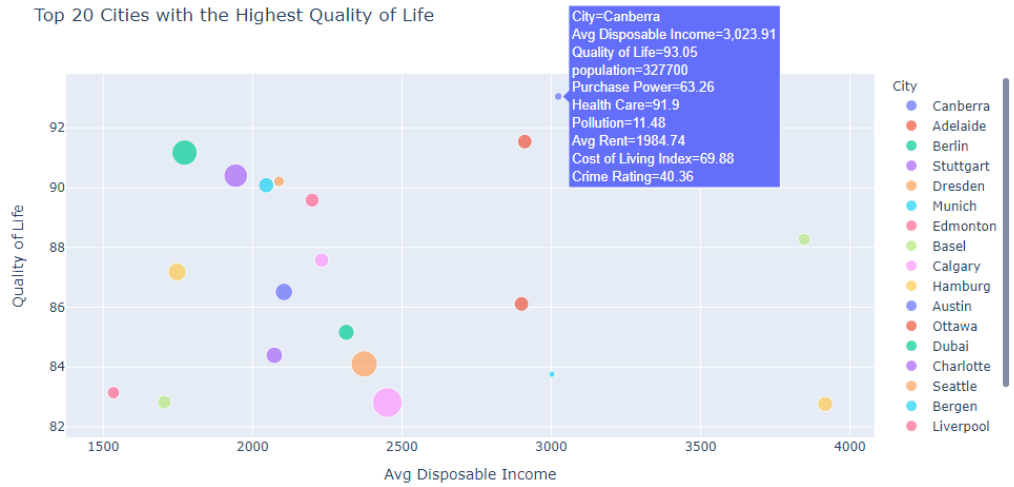
TAMU Datathon

Results - Visualization and Exploration

✓ **Quality of Life**



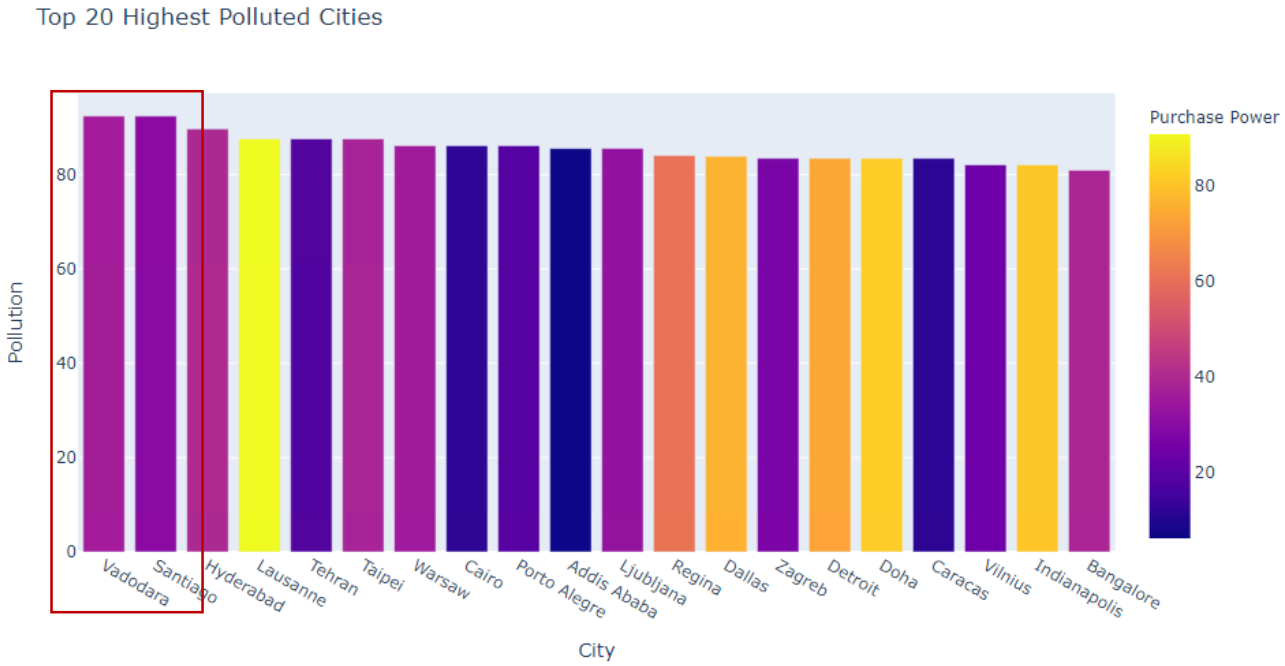
Quality of life vs Avg disposable income



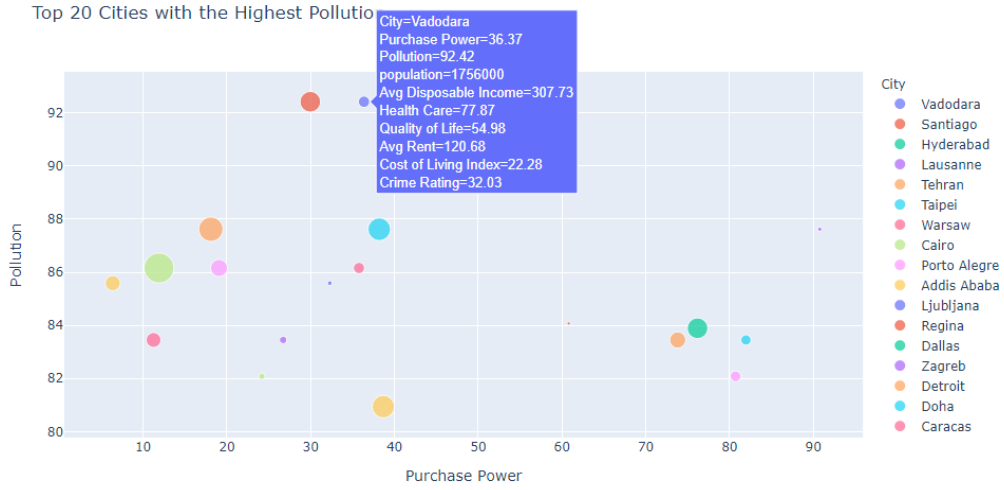
TAMU Datathon

Results - Visualization and Exploration

✓ **Pollution**

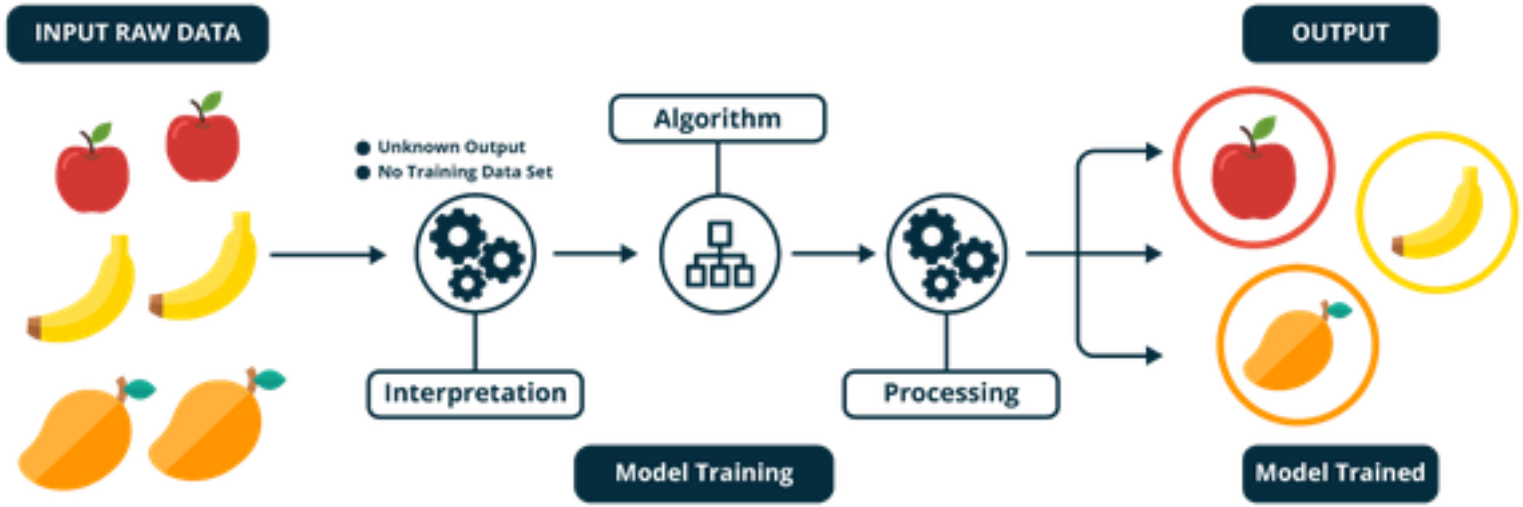


Pollution vs Purchase power



TAMU Datathon

Unsupervised Learning



- Unknown Output
- Pattern

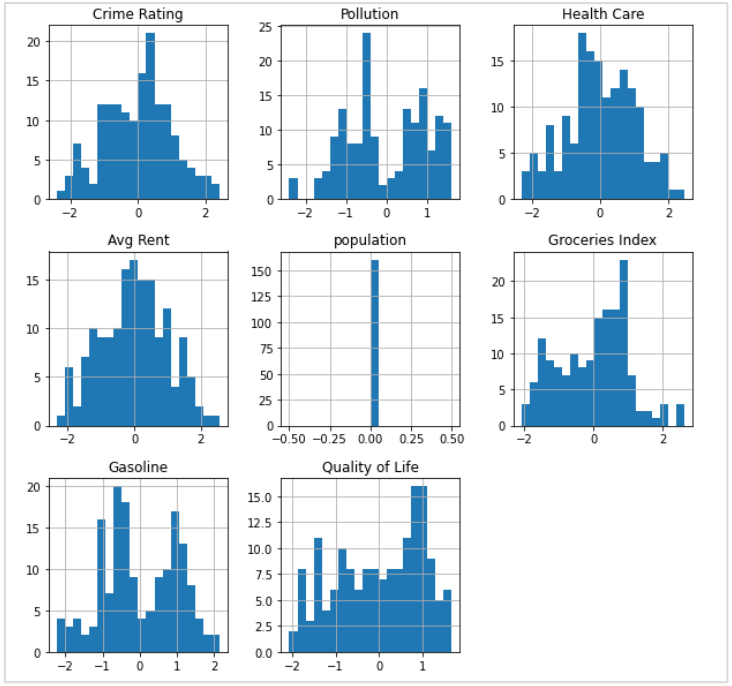
✓ K-Means clustering

✓ Agglomerative clustering

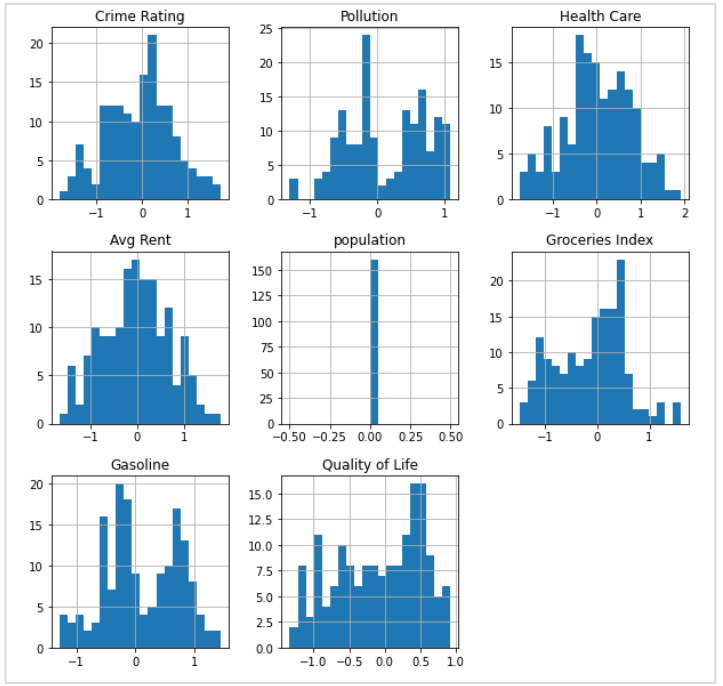
Unsupervised Learning - Preprocessing

Correlation Matrix (Collinearity)

PowerTransform (Gaussian distribution)

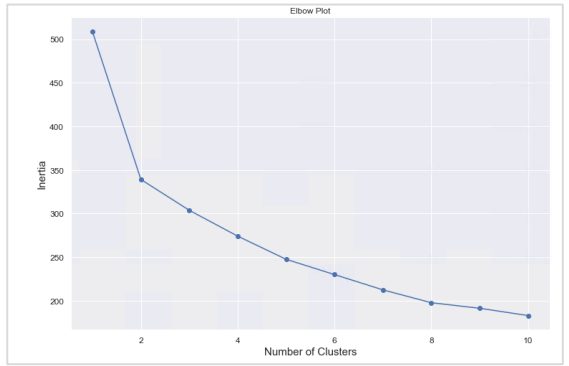


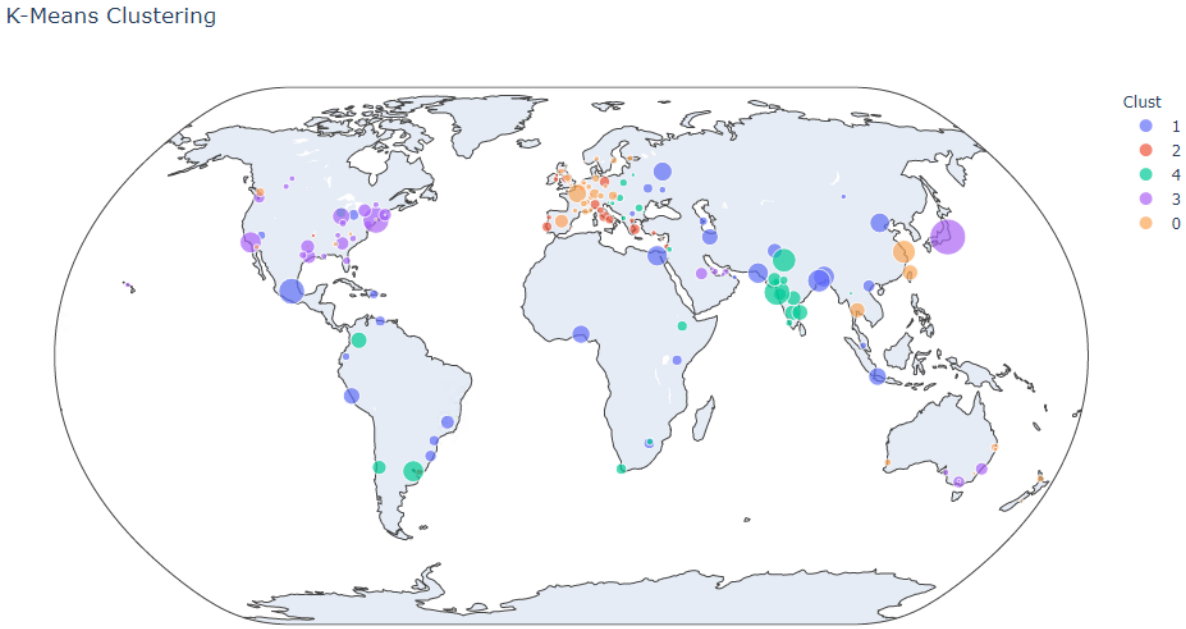
RobustScaler (Scaling)

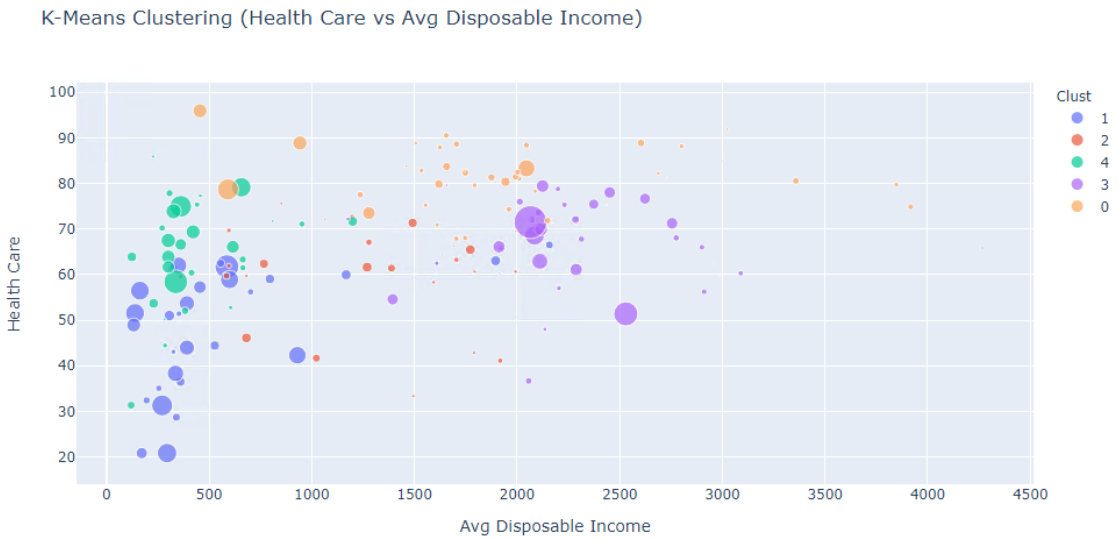


TAMU Datathon

Unsupervised Learning – KMeans clustering

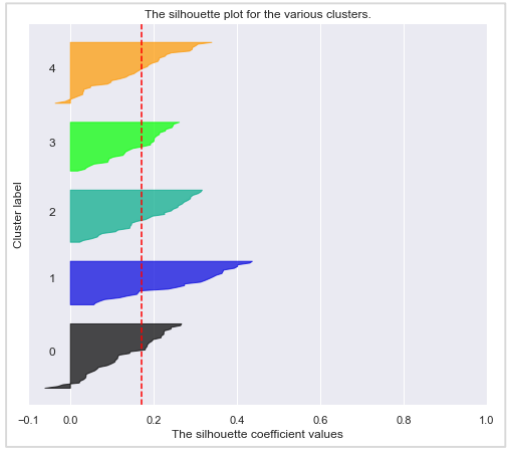


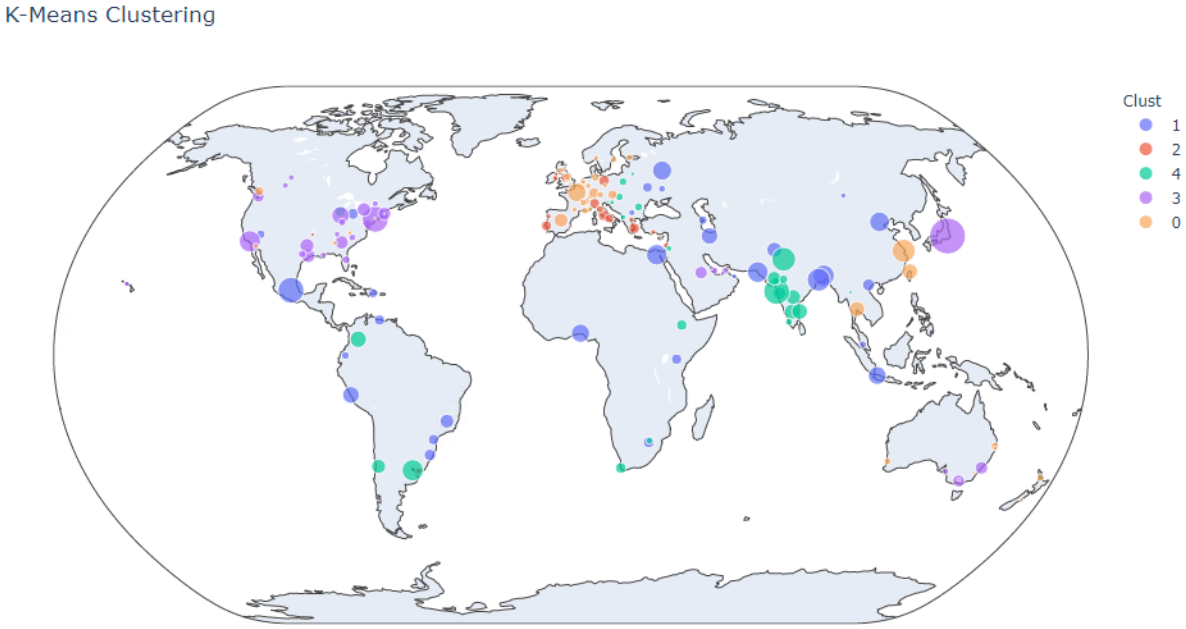


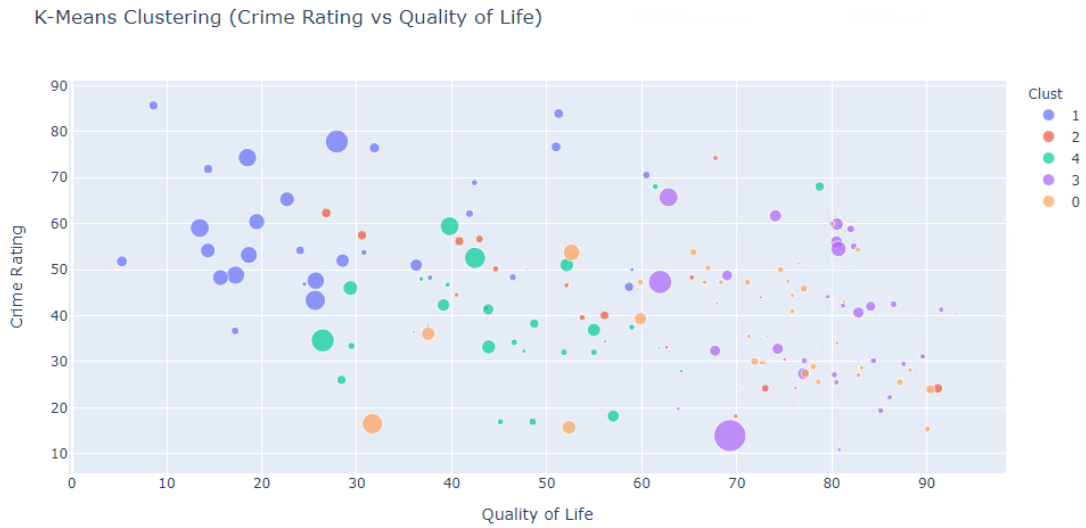


TAMU Datathon

Unsupervised Learning – KMeans clustering

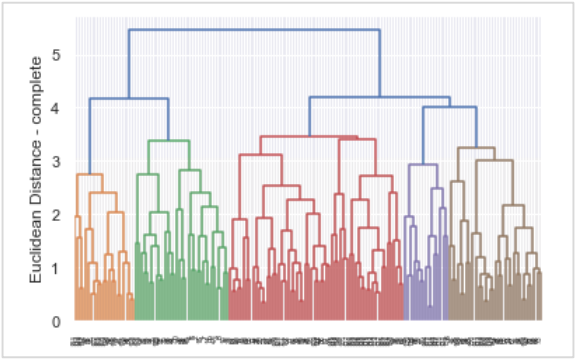


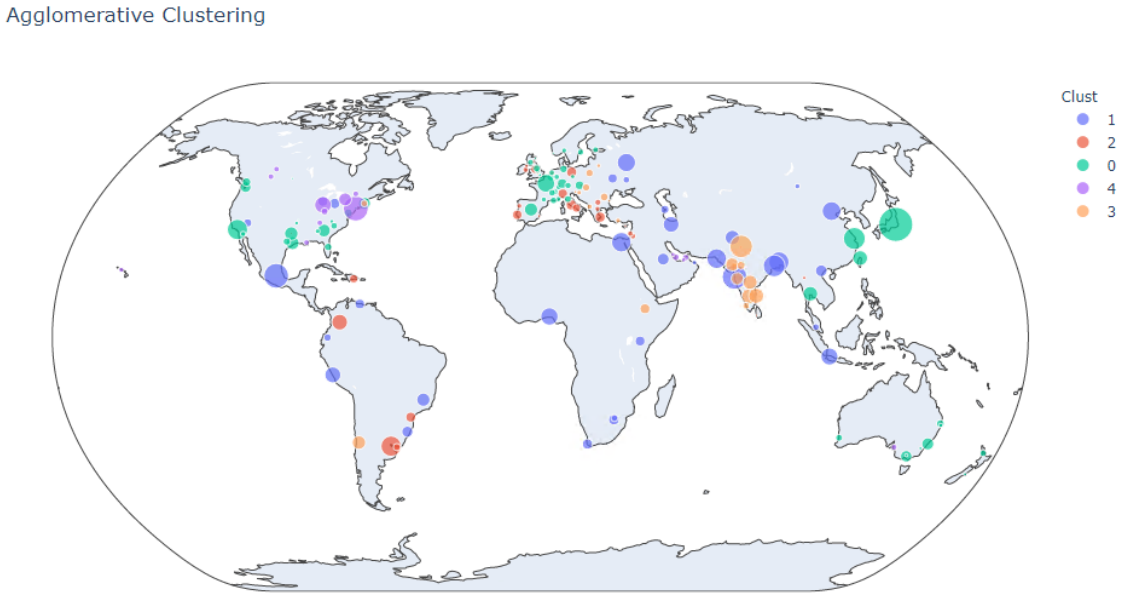


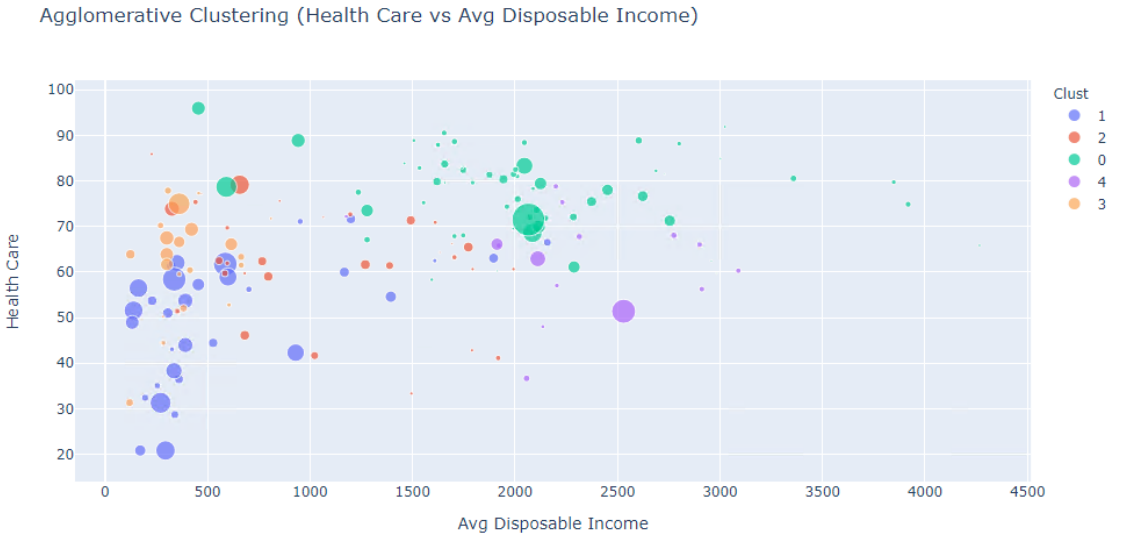


TAMU Datathon

Unsupervised Learning – Agglomerative clustering

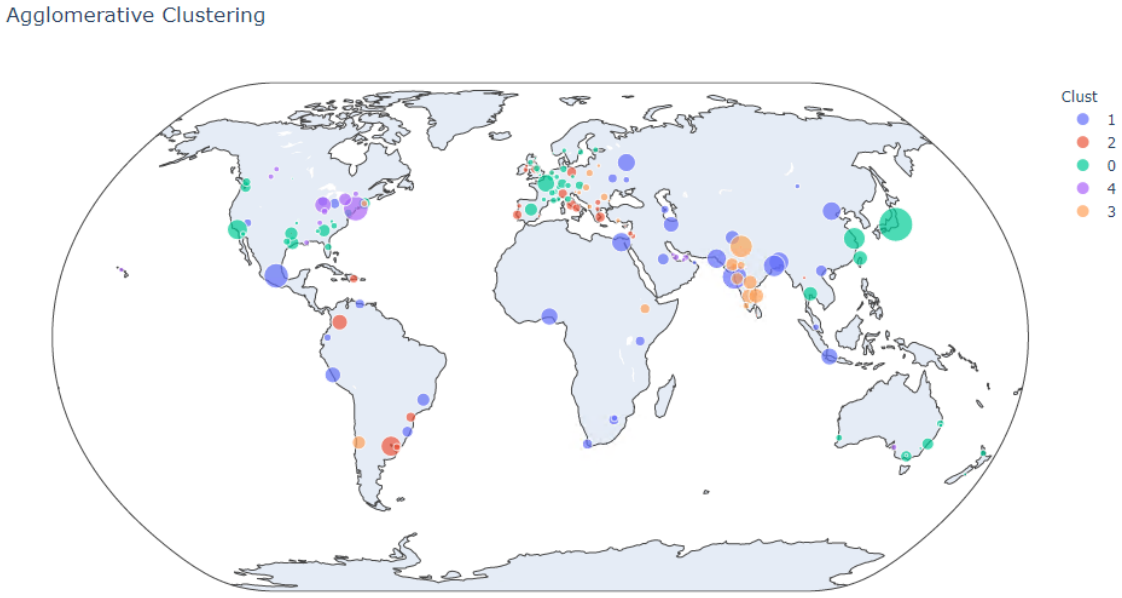






TAMU Datathon

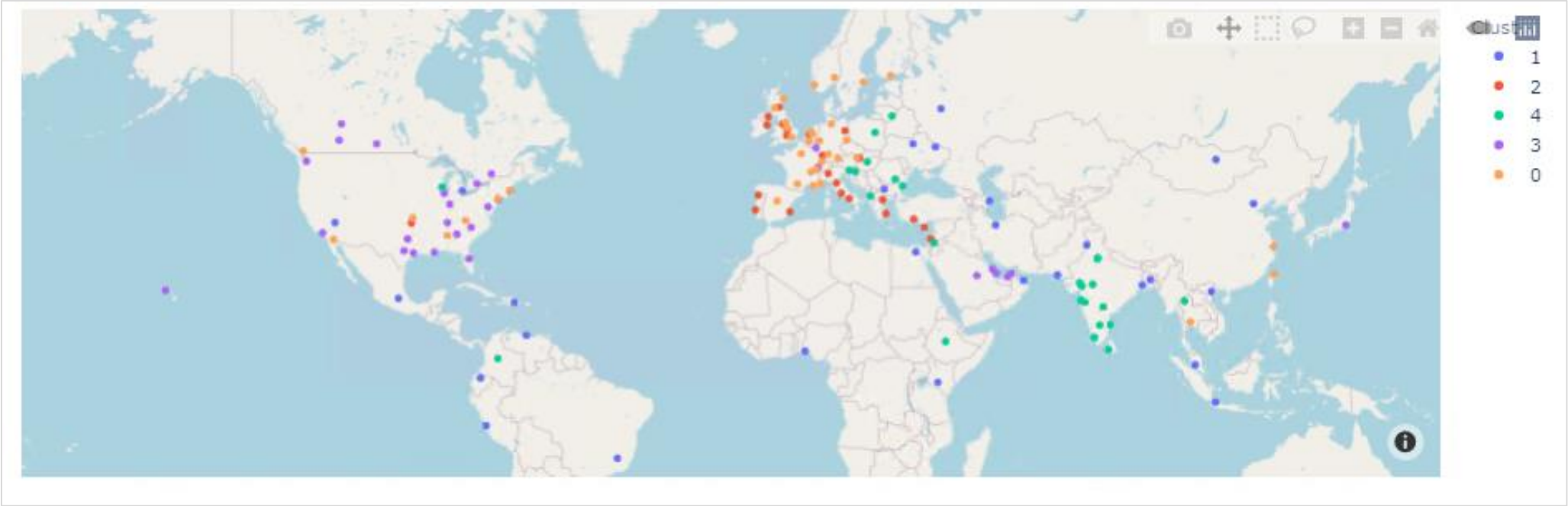
Unsupervised Learning – Agglomerative clustering



TAMU Datathon

Unsupervised Learning

City search tool



TAMU Datathon