EDA of cities across the Globe

A guide: taking into account your personal preferences
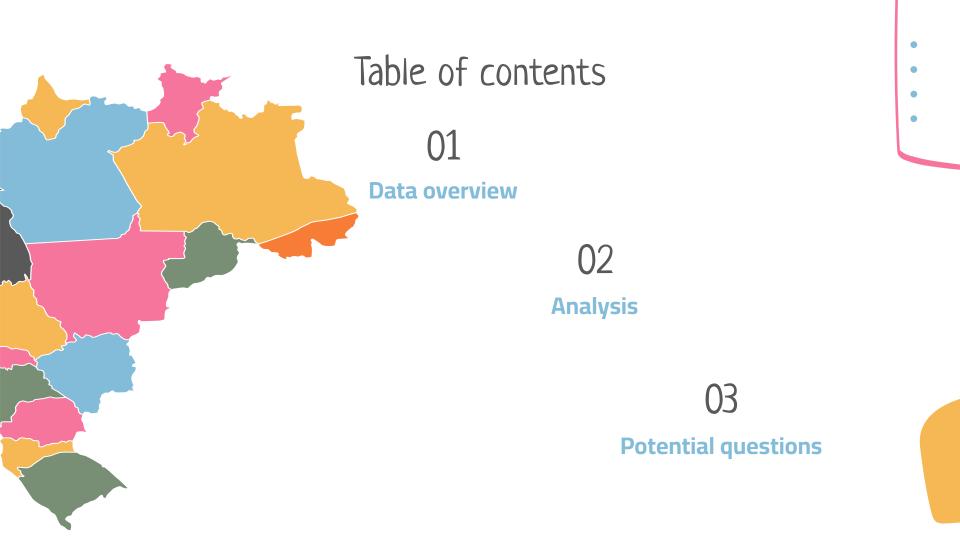
Group – 2

By the combined efforts of:

1. Orish Jindal – ojindal@ucsd.edu
2. Jainish Chauhan – jchauhan@ucsd.edu
3. Zhuoming Zhang - zhz056@ucsd.edu
4. Boyu Chen - boc002@ucsd.edu
5. George Liu - g9liu@ucsd.edu
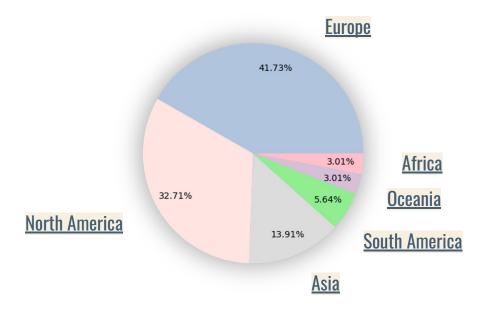6. Sangeetha Viswanathan-Sakthivel - svsakthivel@ucsd.edu

Table of contents

01
Data overview

02
Analysis

03
Potential questions

Data Overview

Dataset: Quality of life in 266 cities, based on 17 factors.

Cities from 6 continents:

Europe 41.73%

Africa 3.01%

Oceania 3.01%

South America 5.64%

Asia 13.91%

North America 32.71%

Analysis

Average scores of features across all cities in the data

Analysis

Average scores of features across all cities in the data
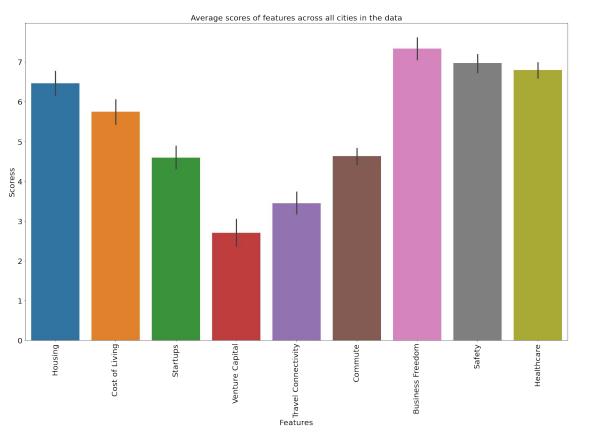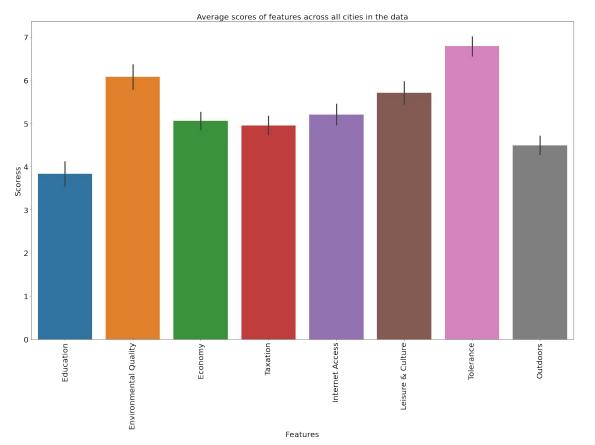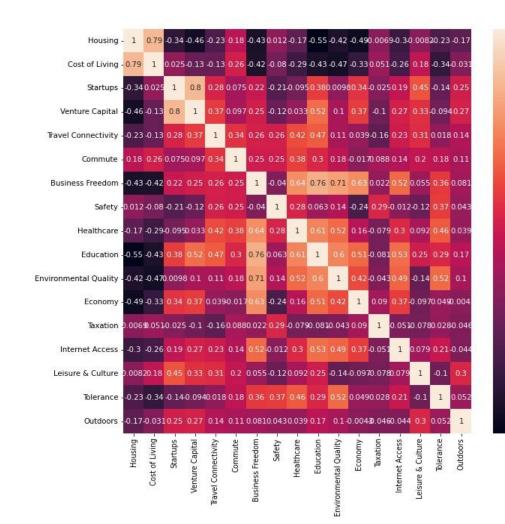
Analysis

Some interesting and relatively unforthcoming relations from heatmap:

- Education-Business freedom
- Business freedom - Environment quality
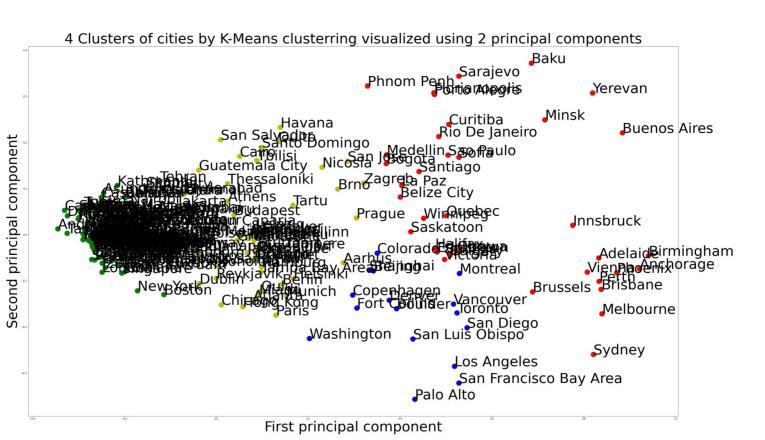- Travel connectivity - Economy
- Cost of living - Taxation

Interesting questions to ask...

1. What cities can be combined into one group and what features are leading to this clustering?

2. What will be the best choice of users with a specific preference?

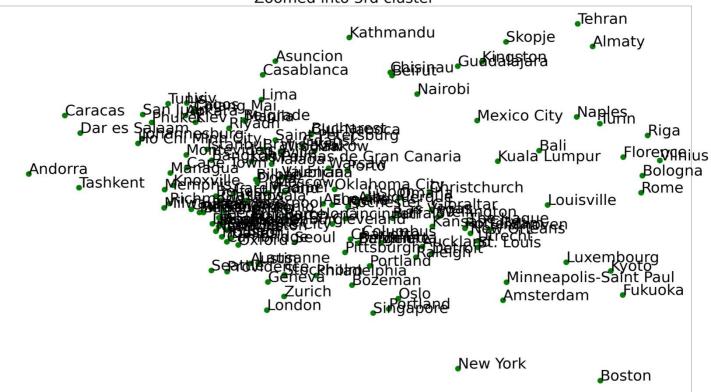3. How different continents are performing in different categories?

Which cities combine using clustering??

4 Clusters of cities by K-Means clusterring visualized using 2 principal components

Which cities combine using clustering??

Zoomed into 3rd cluster

From K-means clustering

- Blue dots:

  - Weather (coastal cities)

  -Tolerance (high)

  -Business Freedom (high)

- Red:

  -Housing (High rating)

  -Commute (medium)
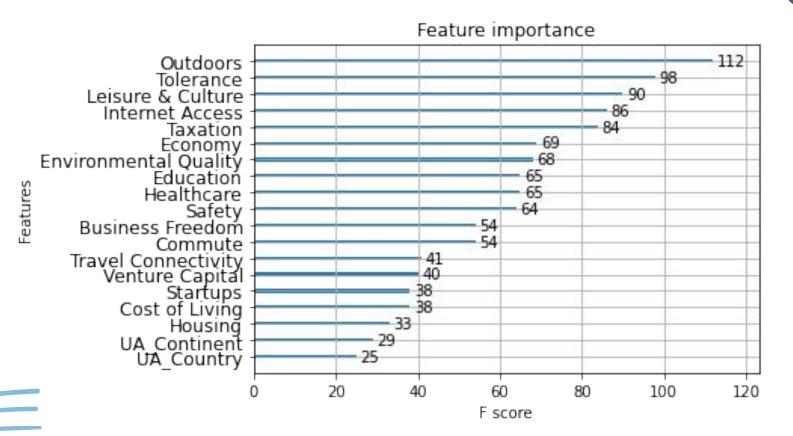
  -Education (Low)

- Light green:

  - Housing (High)

  - Venture Capital (All low)

  - Education (All low)

- Dark green:

  No feature is explicitly matching but a combined effect of all the features is playing a role in clustering the cities of this category together.

Which features play key role?

Feature importance

Feature	F score
Outdoors	112
Tolerance	98
Leisure & Culture	90
Internet Access	86
Taxation	84
Economy	69
Environmental Quality	68
Education	65
Healthcare	65
Safety	64
Business Freedom	54
Commute	54
Travel Connectivity	41
Venture Capital	40
Startups	38
Cost of Living	38
Housing	33
UA_Continent	29
UA_Country	25

Q2: City recommendation

Add content about how we have recommended cities…
        Explain both of the pathways that we have taken (manual and auto combining of features into categories)
What recommendations we are getting for some examples of preferences…
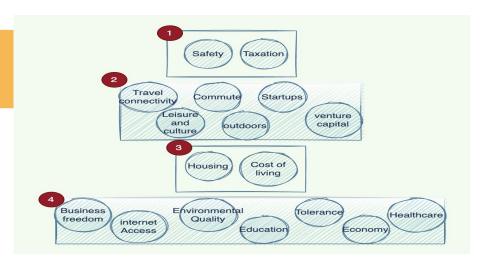What feature relations are helping in getting those rankings
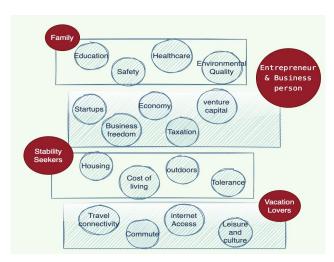What inferences can we draw from it…
Add relevant graphs

City Recommendation

Categorizing features: Since a lot of features are available, they are grouped together in order to cover a wider range of users based on their preferences

Using correlation between features

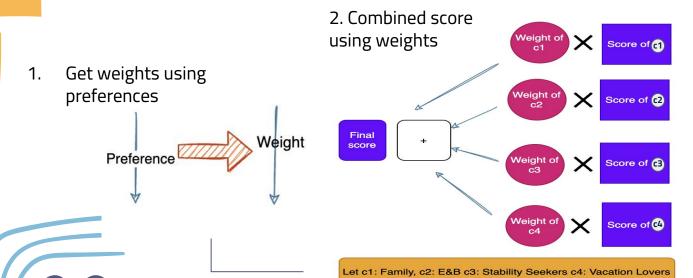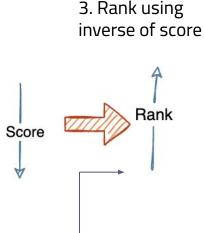Using domain knowledge from user's point of view

City Recommendation

Ranking algorithm (Within category):

Adjust mean value of each column to 5
Take an average of columns within each category

Ranking algorithm (Cross category):

2. Combined score
using weights

1. Get weights using
preferences

Preference → Weight

Weight of c1 × Score of c1

Weight of c2 × Score of c2

Final score + Weight of c3 × Score of c3

Weight of c4 × Score of c4

3. Rank using
inverse of score

Score → Rank

Let c1: Family, c2: E&B c3: Stability Seekers c4: Vacation Lovers

Scores of cities in all categories
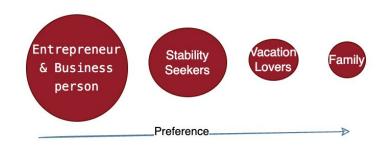
	Vacation Lovers	Entrepreneur & Business person	Stability Seeker	Family
0	4.99	6.20	5.69	8.89
1	5.16	4.29	5.86	8.01
2	4.90	4.58	5.33	3.49
3	4.20	4.49	8.21	3.12
4	6.91	7.89	4.39	8.36
...	...	...	...	...
261	5.20	4.39	5.59	5.95
262	4.07	4.85	5.56	5.73
263	4.19	4.40	9.99	3.40
264	4.23	4.57	7.53	6.82
265	6.71	7.39	2.91	9.43

Q2 City Recommendation

Example:

Entrepreneur & Business person

Stability Seekers

Vacation Lovers

Family

Preference

Final Recommendation

	UA_Name	Vacation Lovers	Entrepreneur & Business person	Stability Seeker	Family	score
2	New York	0.958592	1.000000	0.308786	0.681682	7.525223
1	San Francisco Bay Area	0.989648	0.723280	0.487080	0.729730	7.063385
3	Boston	0.944099	0.657394	0.506460	1.000000	7.037154
5	Chicago	0.857143	0.795022	0.510336	0.567568	6.992949
6	San Diego	0.832298	0.631040	0.702842	0.675676	6.972957

Visualize US Top Cities for Each Category

Top 10 Cities for Family

Top 10 Cities for Stability Seeker

Visualize US Top Cities for Each Category

Top 10 Cities for Entrepreneur
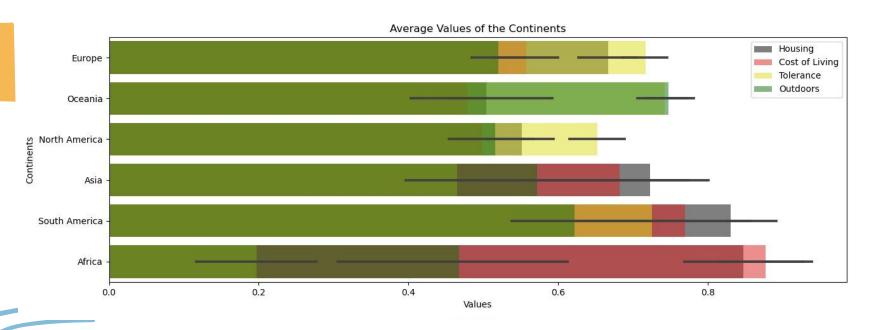
Top 10 Cities for Vacation Lovers

Which continent has which type of cities?
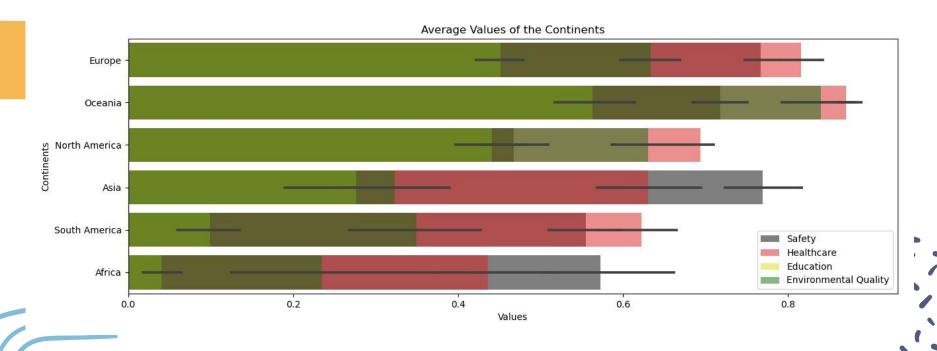
Stability Seeker category: Africa is dominant

Average Values of the Continents

Which continent has which type of cities?
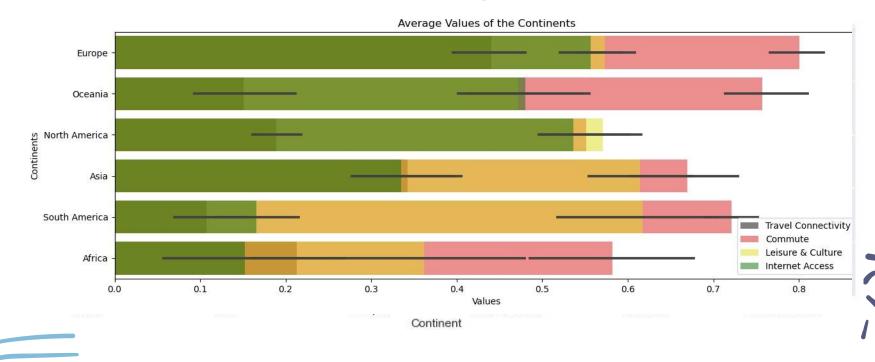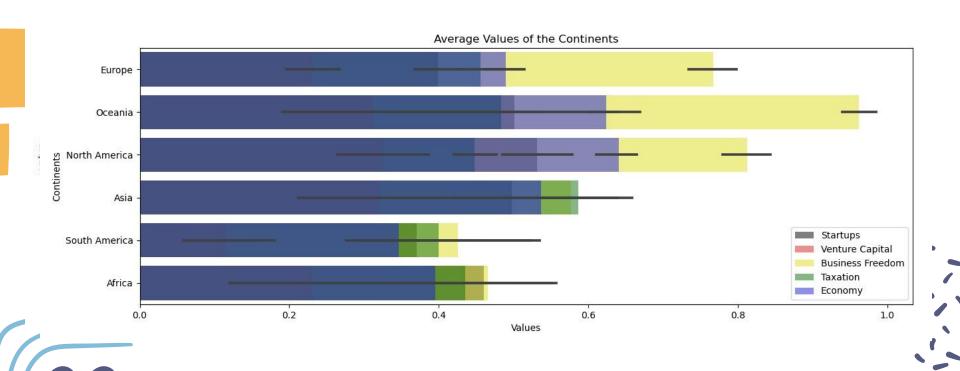
Family category: Oceania is dominant

Average Values of the Continents

Which continent has which type of cities?
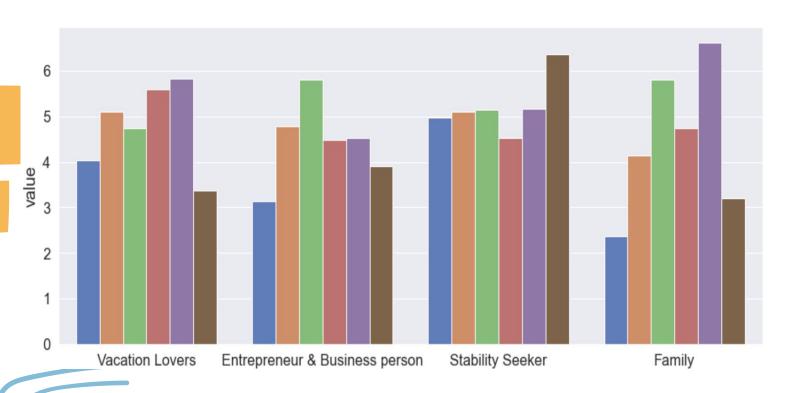
Entrepreneurs and Business Persons category: : Oceania is dominant

Average Values of the Continents

Which continent has which type of cities?

Vacation Lover category: Europe is dominant

Average Values of the Continents

Which continent has which type of cities?
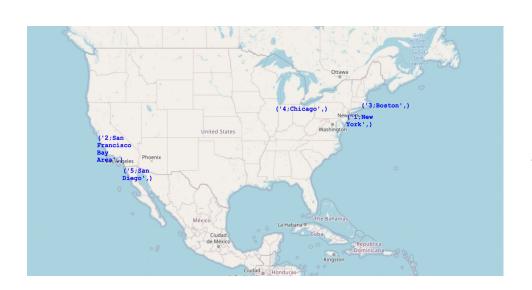
Thanks for listening!

Group – 2

Mark Top Cities on Map

1. New York
2. Boston
3. Los Angeles
4. Chicago
5. San Diego
6. Seattle
7. Austin
8. Salt Lake City
9. Denver
10. Atlanta

Preference: Family > Stability Seeker > Entrepreneurs > Vacation Lovers

Mark Top Cities on Map

1. Raleigh
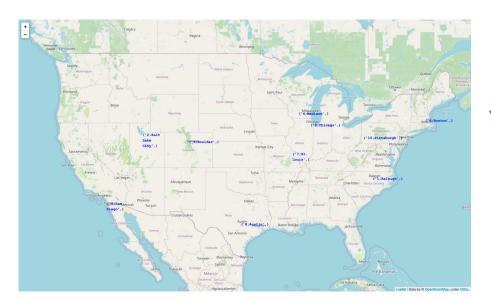2. Salt Lake City
3. Boulder
4. Madison
5. San Diego
6. Boston
7. St. Louis
8. Austin
9. Chicago
10. Pittsburgh

Preference: Category 4 > Category 3 > Category 2 > Category 1