Where to live in Vancouver?

Data Science Capstone Project

Business Problem

- Vancouver (BC, Canada) is a very attractive destination for both immigrants and tourists

- Problem: Decision of where to live or where to stay can be quite challenging

- Solution: A data model that uses location data from Foursquare to rank and recommend the neighborhoods in Vancouver that best fit to each specific preference of immigrants and tourists

- Stakeholders:
    - Immigrants who want to live in Vancouver
    - Tourists who want to visit Vancouver
    - Real estate agents who want to give recommendations to their clients
    - Real estate renting platforms such as Airbnb to help their users searching for neighborhoods that meet their references

Data

- A list of all 22 neighborhoods (or areas) in Vancouver

- Latitude and longitude of each neighborhood

- Foursquare venues data retrieved by calling Foursquare APIs

- Foursquare venue category hierarchy and ID

Methodology

- Get all venues within or around each Vancouver neighborhood and group them into 9 criteria:
    - Entertainment
    - Education
    - Food
    - Nightlife
    - Outdoor
    - Health
    - Religion
    - Shop
    - Transportation

- Users can set their specific preferences by giving weight to each criteria

- The data model will then rank the neighborhoods basing on the specified preferences

- Users can use this ranking as recommendation on where they should live or stay in Vancouver basing on their individual references

Data Analysis Part 1

- Retrieve the list of all neighborhoods in Vancouver and get their coordinates
- Use Python library geopy to get the data

4	Grandview-Woodland	49.275849	-123.066934
5	Hastings-Sunrise	49.277830	-123.040005
6	Kensington-Cedar Cottage	49.246790	-123.073475
7	Kerrisdale	49.220985	-123.159548
8	Killarney	49.218012	-123.037115
9	Kitsilano	49.269410	-123.155267
10	Marpole	49.209223	-123.136150
11	Mount Pleasant	49.264048	-123.096249
12	Oakridge	49.226615	-123.122943
13	Renfrew-Collingwood	49.248577	-123.040179
14	Riley Park	49.244854	-123.103035
15	Shaughnessy	49.246305	-123.138405
16	South Cambie	49.246464	-123.121603
17	Strathcona	49.277693	-123.088539
18	Sunset	49.219093	-123.091665
19	Victoria-Fraserview	49.218980	-123.063816

Data Analysis Part 2

- Get Foursquare categories ID(s) corresponding to each of the 9 criteria

- Get Foursquare data for each neighborhood using /venues/explore API calls

	Neighborhood	Criteria	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Arbutus Ridge	Entertainment	Dunbar Theater	49.245613	-123.185428	Indie Movie Theater
1	Arbutus Ridge	Entertainment	Dance Co	49.248822	-123.154979	Dance Studio
2	Arbutus Ridge	Education	Point Grey Secondary	49.237441	-123.153967	School
3	Arbutus Ridge	Education	Carnarvon Community School	49.256532	-123.173862	School
4	Arbutus Ridge	Education	St. John's School	49.262445	-123.153701	School

Data Analysis Part 3

Create a summary dataframe to make it easier to calculate score and rank the neighborhoods later

0	Arbutus Ridge	2	14	72	4	28	21	8	43	11
1	Downtown	70	93	249	169	148	61	20	184	111
2	Dunbar-Southlands	7	5	31	5	16	4	5	33	4
3	Fairview	44	53	174	60	99	67	12	129	72
4	Grandview-Woodland	11	27	118	30	46	26	9	71	31
5	Hastings-Sunrise	8	16	79	12	31	14	10	50	26
6	Kensington-Cedar Cottage	5	19	150	6	35	29	9	39	46
7	Kerrisdale	5	5	59	5	18	9	6	57	15
8	Killarney	3	8	31	6	17	9	4	43	9
9	Kitsilano	33	25	155	36	86	34	5	119	55
10	Marpole	8	4	63	6	21	7	4	76	30
11	Mount Pleasant	36	49	184	56	79	43	22	109	66
12	Oakridge	7	26	58	5	30	21	15	70	26
13	Renfrew-Collingwood	3	12	90	12	31	17	8	63	31
14	Riley Park	9	25	129	19	49	48	19	78	44
15	Shaughnessy	8	26	111	7	37	25	15	72	33
16	South Cambie	12	29	136	16	50	52	18	87	42
17	Strathcona	36	75	162	74	96	35	12	85	69
18	Sunset	3	25	78	6	18	14	10	48	29
19	Victoria-Fraserview	2	10	55	5	11	15	6	49	17

|	Neighborhood	Entertainment	Education	Food	Nightlife	Outdoor	Healthcare	Religion	Shop	Transportation	Score
1	Downtown	70	93	249	169	148	61	20	184	111	1105
20	West End	65	94	221	164	142	68	16	191	113	1074
3	Fairview	44	53	174	60	99	67	12	129	72	710
11	Mount Pleasant	36	49	184	56	79	43	22	109	66	644
17	Strathcona	36	75	162	74	96	35	12	85	69	644
9	Kitsilano	33	25	155	36	86	34	5	119	55	548
16	South Cambie	12	29	136	16	50	52	18	87	42	442
14	Riley Park	9	25	129	19	49	48	19	78	44	420
4	Grandview-Woodland	11	27	118	30	46	26	9	71	31	369
6	Kensington-Cedar Cottage	5	19	150	6	35	29	9	39	46	338
15	Shaughnessy	8	26	111	7	37	25	15	72	33	334
13	Renfrew-Collingwood	3	12	90	12	31	17	8	63	31	267
12	Oakridge	7	26	58	5	30	21	15	70	26	258
5	Hastings-Sunrise	8	16	79	12	31	14	10	50	26	246
18	Sunset	3	25	78	6	18	14	10	48	29	231
10	Marpole	8	4	63	6	21	7	4	76	30	219
0	Arbutus Ridge	2	14	72	4	28	21	8	43	11	203
7	Kerrisdale	5	5	59	5	18	9	6	57	15	179
19	Victoria-Fraserview	2	10	55	5	11	15	6	49	17	170
21	West Point Grey	5	11	58	6	29	8	5	41	7	170
8	Killarney	3	8	31	6	17	9	4	43	9	130

Data Analysis Part 4

Take user's preference, calculate score and rank the neighborhoods
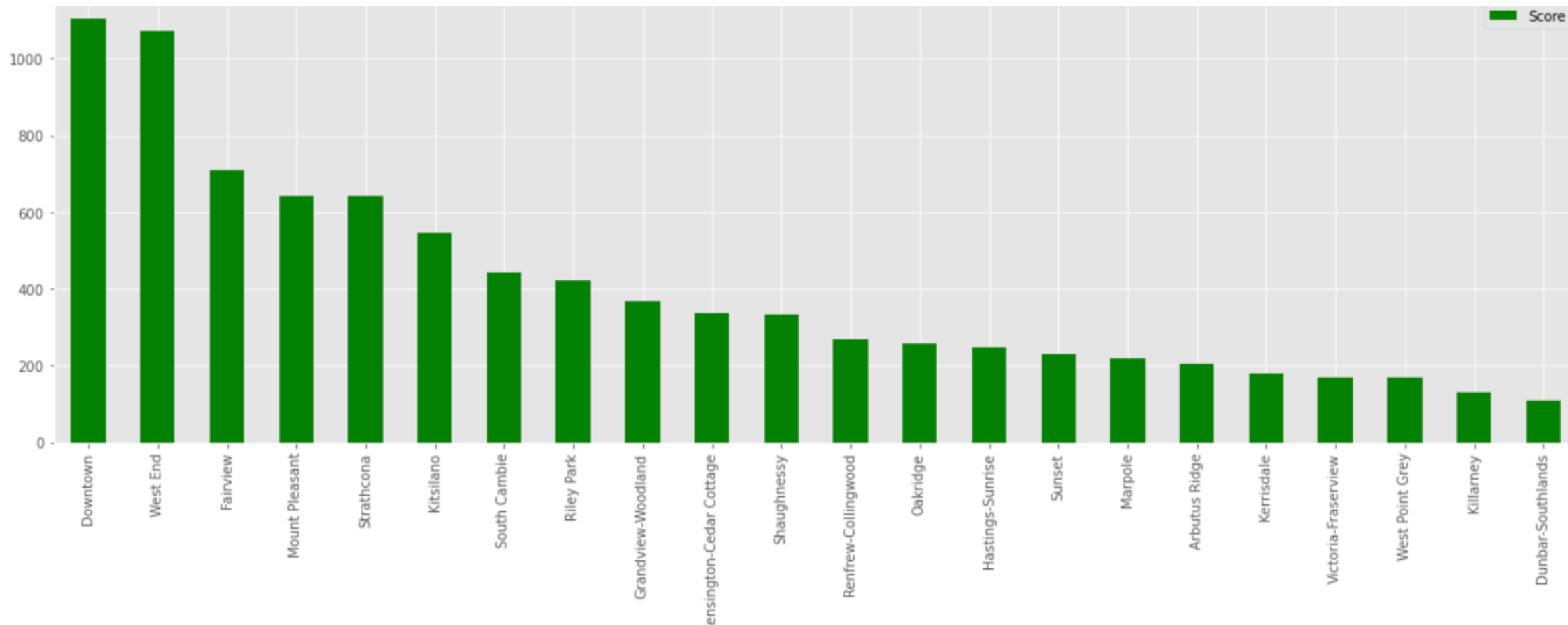
Results | Preference 1: A user who give the same weight to all criteria
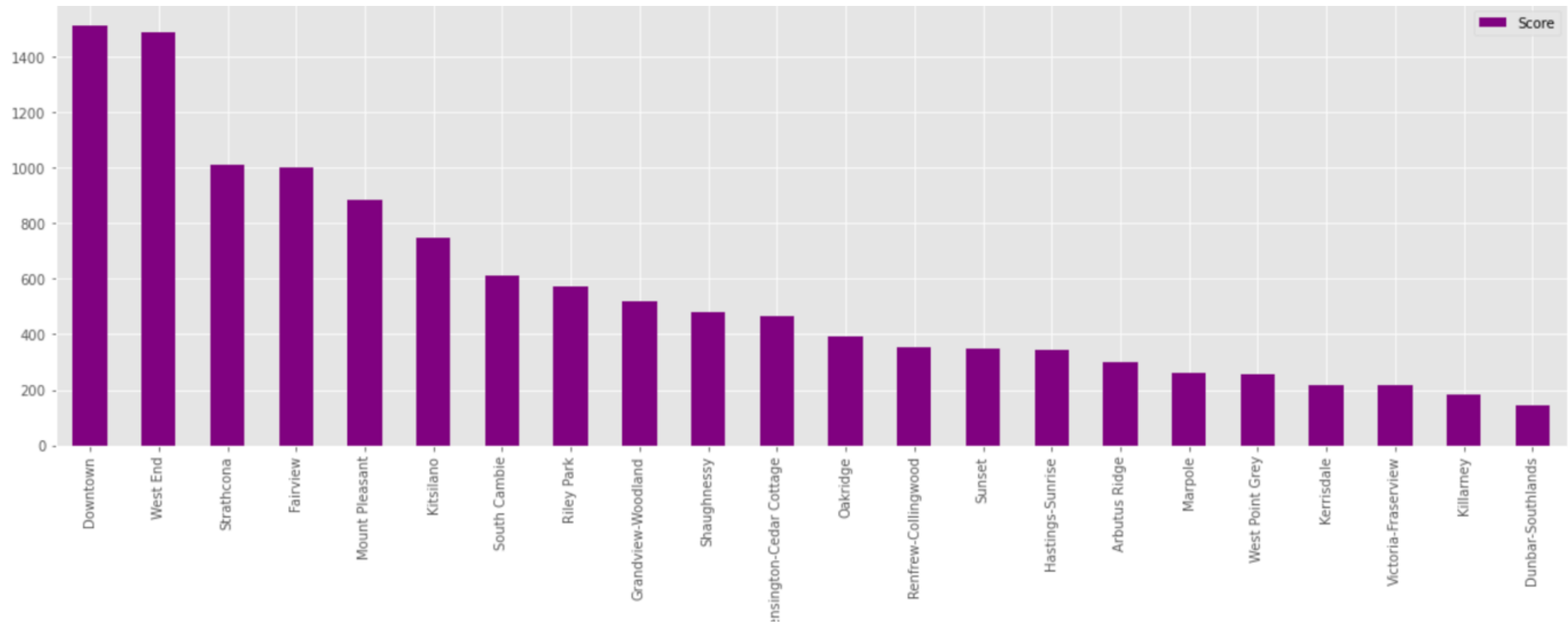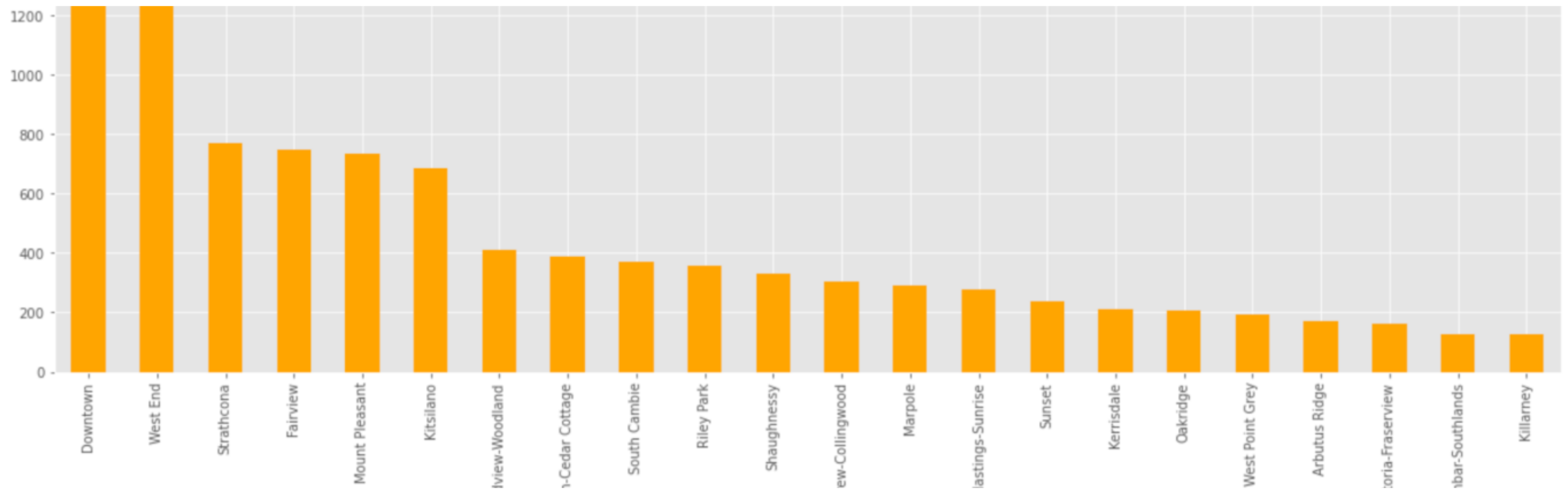
Results | Preference 2: A family with kids that give high weight to Education, Outdoors, Food and Shop

Results

- Preference 3: A young tourist couples who give high weights to Entertainment and Nightlife but doesn't want to stay near spiritual centers and hospitals.

Discussion

- Weight for each criteria can also be set to 0 or even to a negative number
- Venues in Vancouver are not equally distributed across the neighborhoods. Most of them are in Downtown and West End => In most cases, these neighborhoods are in the top 2, no matter what the user's preference is
- In most cases, these neighborhoods are in the top 2, no matter what the user's preference is
- Limitation: Quality of the venues are not taken into account

Conclusion

- This data model can be used by immigrants, tourists and real estate agents to rank the 22 neighborhoods in Vancouver basing on specific "preferences". Therefore, it can help the stakeholders to make decision on where to live or stay in Vancouver basing on their own needs.

- Future improvements:
  - Take into account quality of the venues (ratings, reviews, check-ins, etc…)
  - A web application to expose the model to the users