

**Data Science Capstone Project**

-

**The best city for a new business  
(Movie Theater)**

# **Aim of the project**

## **'Best city for a new business'**

- **Support business founder with location selection**
  - A new movie theater to be established
  - A suited location has to be identified in the US
- **Provide solid information for a profound decision**
  - Conduct data analysis aiming at a recommendation
- **Provide a template for similar analyses**

# Project Background

- **Key addressee aims at opening a new movie theater**
  - Needs support to evaluate the best-suited location
- **Secondary goal is the expansion to multiple theaters**
  - Among the best-suited city for a single theater
  - Throughout the country (not direct scope of this project)
- Boundary Condition: **The business has to be in the US**

# High-Level Approach

- **Acquire information on US cities to investigate their goodness to host the new business**
  - Location, population, zip codes, venues
- **Select a good sample of cities for a detailed assessment**
- **Compare the cities and their neighborhoods to identify the best-suited for the business**
  - Based on comparison of movie theaters with coffee shops and bars
- **Select the best-fitting city based on the neighborhood evaluation**

# The proper city sample

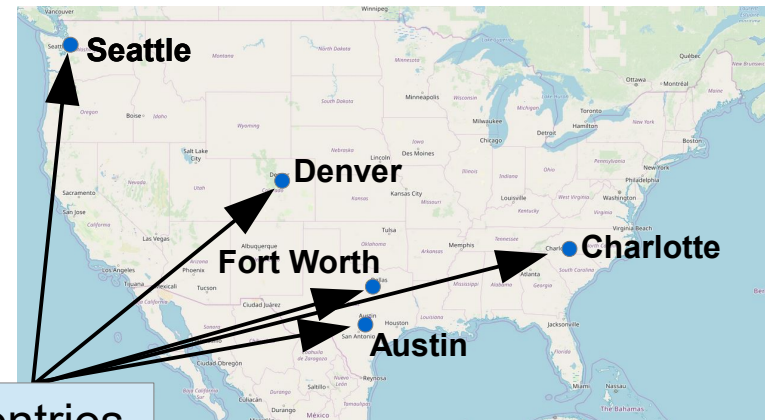
- A table of cities with more than 500,000 residents lists 35 cities
- Five of these cities showing the highest population growth rate are selected for further investigation

Num	City	2018 estimate	2010 census	Change [%]	Area [km <sup>2</sup> ]	Pop. Density [/km <sup>2</sup> ]	Latitude	Longitude
17	18	Seattle	744955	608660	22.39	217.0	3,245	47.60357 -122.32945
10	11	Austin	964254	790390	22.00	809.9	1,170	30.26759 -97.74299
12	13	Fort Worth	895008	741206	20.75	888.1	962	32.75095 -97.33086
18	19	Denver	716492	600158	19.38	397.0	1,746	39.74001 -104.99202
15	16	Charlotte	872498	731424	19.29	791.0	1,064	35.22286 -80.83796
19	20	Washington	702455	601723	16.74	158.2	4,304	38.89037 -77.03196
34	35	Mesa	508958	439041	15.92	357.2	1,357	33.41704 -111.83146
6	7	San Antonio	1532233	1327407	15.43	1,194.0	1,250	29.42458 -98.49461
4	5	Phoenix	1660272	1445632	14.85	1,340.6	1,200	33.44825 -112.07580
13	14	Columbus	892533	787033	13.40	565.9	1,520	39.96199 -83.00275
20	21	Boston	694583	617594	12.47	125.1	5,381	42.35666 -71.05674
8	9	Dallas	1345047	1197816	12.29	882.9	1,493	32.77815 -96.79540
26	27	Oklahoma City	649021	579999	11.90	1,570.3	407	35.47203 -97.52107
24	25	Portland	653115	583776	11.88	345.8	1,851	45.51179 -122.67563
23	24	Nashville	669053	601222	11.28	1,232.6	536	36.16784 -86.77816
3	4	Houston	2325502	2100263	10.72	1,651.1	1,395	29.76058 -95.36968
27	28	Las Vegas	646444	583756	10.43	348.1	1,818	36.17193 -115.14001
11	12	Jacksonville	903889	821784	9.99	1,935.8	455	30.33147 -81.65622
14	15	San Francisco	883305	805235	9.70	121.5	7,170	37.77712 -122.41964
7	8	San Diego	1425976	1307402	9.07	842.3	1,670	32.71568 -117.16171
35	36	Sacramento	508529	466488	9.01	253.6	1,953	38.57944 -121.49085
9	10	San Jose	1030119	945942	8.90	459.7	2,231	37.33865 -121.88542
33	34	Fresno	530093	494665	7.16	296.3	1,762	36.74084 -119.78552
16	17	Indianapolis	867125	820445	5.69	936.3	914	39.76691 -86.14996
1	2	Los Angeles	3990456	3792621	5.22	1,213.9	3,276	34.05349 -118.24532
21	22	El Paso	682669	649121	5.17	665.1	1,030	31.75916 -106.48749
32	33	Tucson	545975	520116	4.97	597.8	888	32.22155 -110.96976
5	6	Philadelphia	1584138	1526006	3.81	347.6	4,511	39.95222 -75.16218
28	29	Louisville	620118	597337	3.81	682.5	903	38.25489 -85.76666
0	1	New York	8398748	8175133	2.74	780.9	10,933	40.71455 -74.00714
31	32	Albuquerque	560218	545852	2.63	487.4	1,147	35.08423 -106.64905
25	26	Memphis	650618	646889	0.58	822.1	794	35.14976 -90.04925
2	3	Chicago	2705994	2695598	0.39	588.7	4,600	41.88425 -87.63245
30	31	Milwaukee	592025	594833	-0.47	249.2	2,388	43.04200 -87.90687
29	30	Baltimore	602495	620961	-2.97	209.5	2,934	39.29058 -76.60926
22	23	Detroit	672662	713777	-5.76	359.5	1,871	42.33168 -83.04800

35 entries  
> 500,000 res.

Top5 growth rate

5 entries





Num	City	2018 estimate	2010 census	Change [%]	Area [km <sup>2</sup> ]	Pop. Density [/km <sup>2</sup> ]	Latitude	Longitude
0	18	Seattle	744955	608660	22.39	217.0	3,245	47.60357 -122.32945
1	11	Austin	964254	790390	22.00	809.9	1,170	30.26759 -97.74299
2	13	Fort Worth	895008	741206	20.75	888.1	962	32.75095 -97.33086
3	19	Denver	716492	600158	19.38	397.0	1,746	39.74001 -104.99202
4	16	Charlotte	872498	731424	19.29	791.0	1,064	35.22286 -80.83796

# Parameter for the assessment

- Venue information are taken from Foursquare.com:
  - Number of movie theaters, coffee shops and bars
    - # of Coffee shops and bars considered representative for a vivid public life
  - For an area of 1 km radius around each zip code location

Parameter to assess the goodness for hosting the new business:

$$\text{Evaluation index} = (\text{Coffee shops} + \text{bars}) / \text{Movie Theaters}$$

- More coffee shops and bars  Higher evaluation index
- Fewer movie theaters  Higher evaluation index
- A higher index represents a better goodness for the new business!

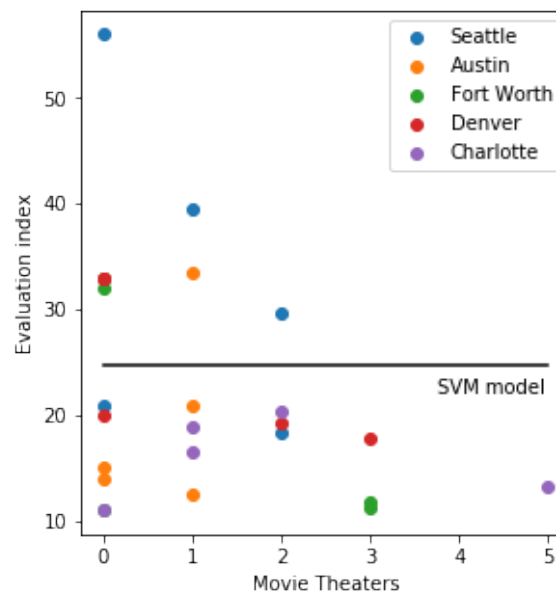
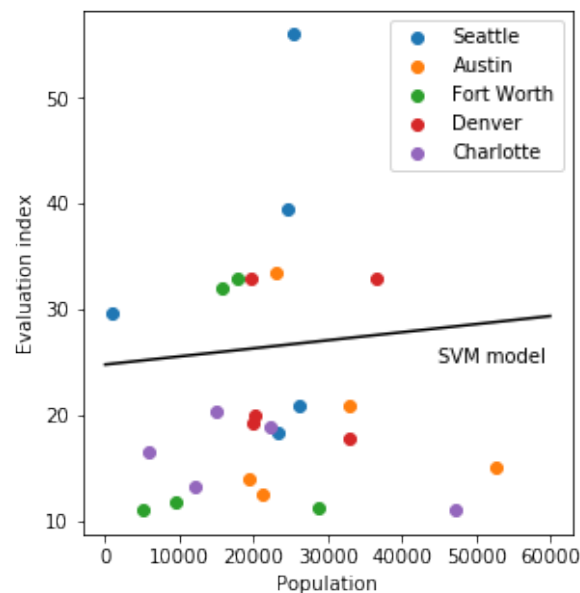
# Calculations

- The evaluation index has been computed for each zip code of the five considered cities
- Assessment/Comparison of the evaluation index has been done for:
  - The five highest indices of each city
    - Scatter plot (slide 8)
  - The distributions of the evaluation indices of each city
    - Boxplot (slide 9)
- A Support Vector Machine has been applied to discriminate the well-suited areas from the less suited
  - Allows confirmation/support of the results
  - Allows a future repetition to assess 'new' neighborhoods



# Results 1

- Evaluation index against population and movie theaters displays Seattle as the best-suited city:
  - The two highest evaluation indices are located in **Seattle** and significantly higher than the rest
  - **Seattle** shows the highest fraction of zip code areas above the SVM limiter

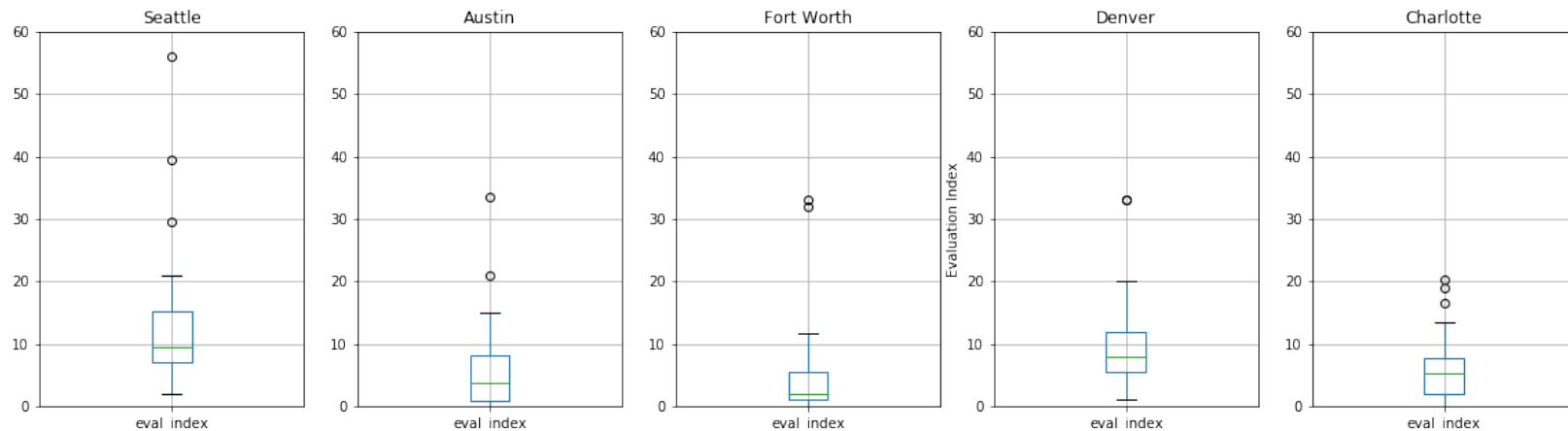


City	Above limiter
Seattle	3/5
Fort Worth	2/5
Denver	2/5
Austin	1/5
Charlotte	0/5



## Results 2

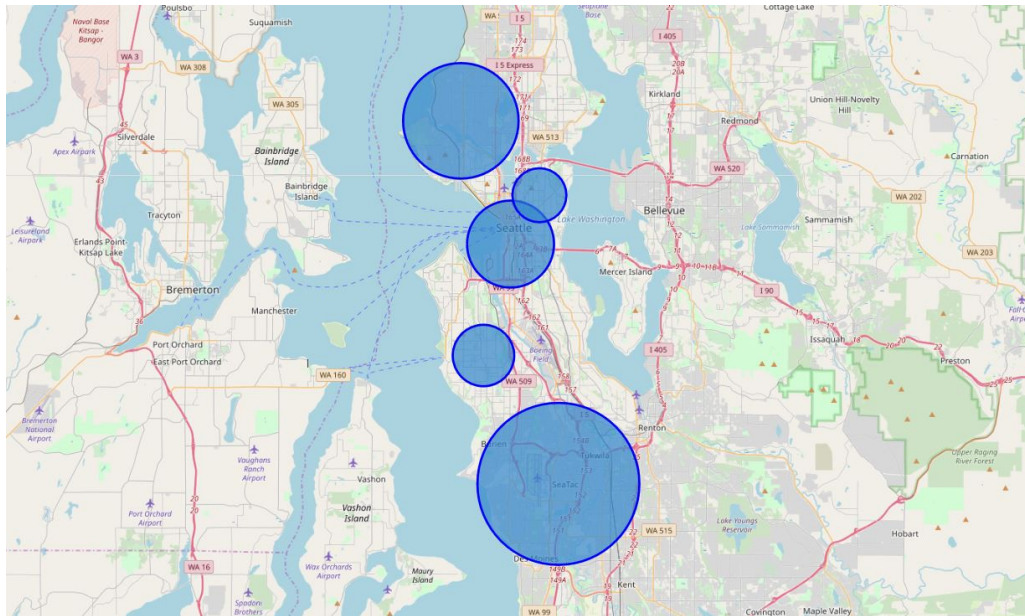
- The boxplot depicts that all relevant parameters are highest in **Seattle**



City	Lo. Quartile	Median	Up. Quartile	3 <sup>rd</sup> Highest	2 <sup>nd</sup> Highest	Highest
Seattle	7.04	9.56	15.13	29.67	39.50	56.00
Austin	0.88	3.75	8.10	15.00	21.00	33.50
Fort Worth	1.00	2.00	5.50	11.75	32.00	33.00
Denver	5.50	8.00	11.97	20.00	33.00	33.00
Charlotte	2.00	5.17	7.71	16.50	19.00	20.33

# Summary and conclusion

- The assessment of the maximum evaluation index favors **Seattle**
- The assessment of the value distribution of the evaluation indices among the considered cities highly favors **Seattle**
- The distribution of the top5 zip code areas in **Seattle** are in supportive for a future expansion



## Recommendations:

- Theater(s) should open in Seattle
- First theater should open at ZIP code 98188
- First theater should open at ZIP code 98107