

The background features a series of concentric circles in light gray, some solid and some dashed, creating a ripple effect. A large red speech bubble is centered on the page, containing the title text in white.

Analysis of tourist arrivals in cities around the world

The background of the slide features several sets of thin, curved lines in a light gray color, some solid and some dashed, creating a sense of motion or a stylized globe. A large red speech bubble is positioned on the left side of the slide.

1. Interest

- City tourist management agencies can be very interested to have a tool that they can use to increase the number of annual tourists and improve the economy of their city.
- The tool can be used to understand which services need to be improved to attract more tourists

2. Data acquisition and cleaning

■ DATA ACQUISITION:

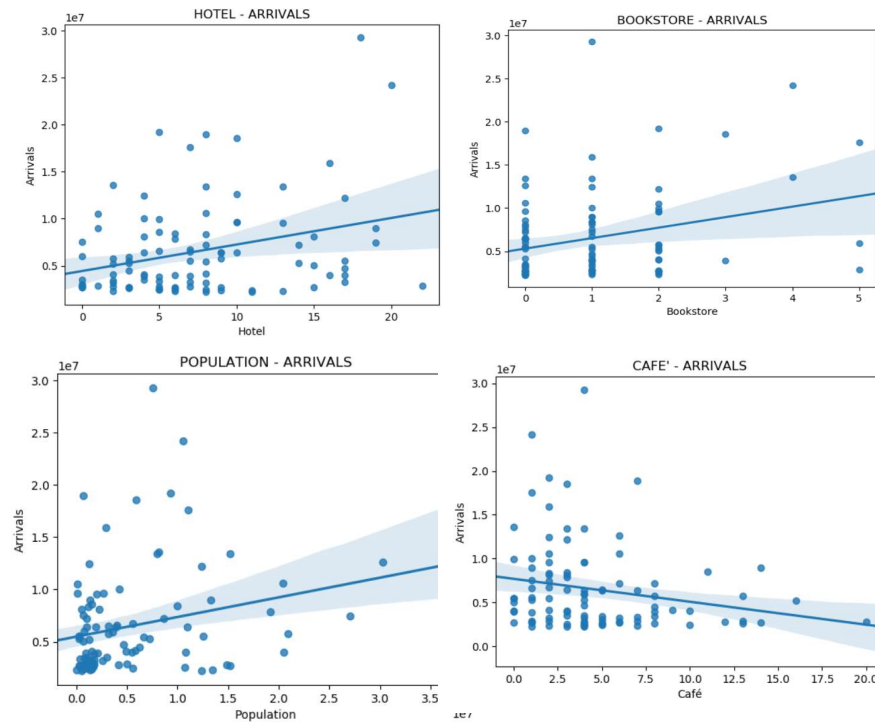
- LIST OF CITIES 1: I use the "List of cities by international visitors" (https://en.wikipedia.org/wiki/List_of_cities_by_international_visitors) classified by the Euromonitor Rank. I'm interested in the "City" and "2018 Arrivals" columns. For every city I will find Latitude and Longitude with geolocator and add these two columns to the DataFrame.
- LIST OF CITIES 2: I use the list of city by population (csv file: <https://worldpopulationreview.com/world-cities>)
- LIST OF CITIES 3: I use the list of city by area (<http://www.citymayors.com/statistics/largest-cities-area-125.html>).
- I merge the three data-frames. The Data frame will have the following columns: City, Population, Area, Arrivals, Latitude, Longitude, Population, Land Area, Arrivals.
- FINAL DATASET: I will add to this dataset the number of venues (restaurant, hotel, etc.) for each city, getting the data from Foursquare.

■ DATA CLEANING AND FEATURES SELECTION:

Data downloaded or scraped from multiple sources were combined into one table. There were a lot of missing values, mainly geographic and demographic data. I have to fill the missing values manually. After data cleaning and feature selection there were 100 samples and 28 features in the data.

3. Data Analysis

PEARSON CORRELATION AND P-VALUE



Pearson correlation and P-value

	Venues	Corr arrivals	p_value arrivals
6	Café	-0.203843	0.0419
5	Burger Joint	-0.179115	0.0746
15	Pizza Place	-0.17905	0.0747
18	Restaurant	-0.137721	0.1718
3	Beach	-0.136569	0.1755
17	Resort	-0.130462	0.1958
8	Coffee Shop	-0.115496	0.2525
22	Supermarket	-0.07001	0.4888
1	Bakery	-0.0614444	0.5437
2	Bar	-0.0459773	0.6497
16	Plaza	-0.0231188	0.8194
1	Land Area	-0.0216751	0.8305
1	Scenic Lookout	-0.0166459	0.8694
1	Theater	-0.0121452	0.9045
1	Ice Cream Shop	-0.0117868	0.9073
1	Historic Site	-0.0102916	0.9191
1	Sandwich Place	-0.00861017	0.9322
1	Wine Bar	0.00172134	0.9864
1	Dessert Shop	0.00972386	0.9235
1	Cocktail Bar	0.0331894	0.7431
1	Park	0.041282	0.6834
1	Art Museum	0.0525708	0.6034
1	Garden	0.17683	0.0784
21	Shopping Mall	0.188032	0.061
25	Population	0.261778	0.0085
4	Bookstore	0.28316	0.0043
12	Hotel	0.30597	0.002

3. Cluster Creation

Cluster	Population	Land Area	Arrivals
High	8.546208e+06	1555.969697	1.196490e+07
Medium	4.610400e+06	1311.151515	4.971482e+06
Low	3.237467e+06	1850.882353	2.695406e+06



4. Predictive Model

■ **DECISION TREE:**

■ Decision tree parameters:

```
DecisionTreeClassifier(class_weight=None, criterion='entropy', max_depth=6,  
                        max_features=None, max_leaf_nodes=20,  
                        min_impurity_decrease=0.0, min_impurity_split=None,  
                        min_samples_leaf=1, min_samples_split=2,  
                        min_weight_fraction_leaf=0.0, presort=False, random_state=None,  
                        splitter='best')
```

■ Model Evaluation: accuracy = 0,89

■ Feature importance

Feature	Ranking
Land Area	0.197321
Restaurant	0.175164
Coffee Shop	0.172012
Café	0.142952
Hotel	0.102380
Plaza	0.050713
Historic Site	0.049559
Theater	0.048597
Art Museum	0.034768
Garden	0.026535

5. Conclusions

Excluding the land area, this study shows that services (hotels, restaurants, etc.) are the structures that have the most influence on tourism numbers. The tourism management agencies they must therefore take into account that more than the attractions of the place (parks, historical sites, etc.) are not enough to increase the number of tourists, but they need to improve the services.