**APPLIED DATA SCIENCE CAPSTONE**

**Final assessment: Tenedor Gaucho**

**By Florencia Paradeda**

**Part 1**

**Introduction**

* + Argentina is a large country in South America. Best known for its most famous soccer player: Lionel Messi, but also for it’s amazing cuisine.
    - The best steak in the world, empanadas and locro are only a few examples of a large variety of unique delicacies.
  + Argentinians are explores. Most Argentinians are descendants of immigrants who pursed a different, better life, so exploration is part of our lives. About 1 million Argentinians living abroad, usually in large cities, and often show the world the wonders of such unique cuisine.
  + Argentina’s capital city, Buenos Aires, is an electric, hectic location with a wonderful architecture, very active nightlife, and a wide offer of both local and international restaurants. It attracts 2 millions of international tourists every year who spend 33% of their travel budget in restaurants.

**Objective**

* The objective of this project is to recommend a location that presents a business opportunity for opening and Argentinian Restaurant.
* This analysis will result of interest for Argentinians living in such cities, or for investors looking for business opportunities.
* A subset of major cities with significant Argentinian presence will be part of this analysis:
  + Barcelona, London, Madrid, Mexico City, New York, Sao Paulo, and Toronto.

**APPLIED DATA SCIENCE CAPSTONE**

**Final assessment: Tenedor Gaucho**

**By Florencia Paradeda**

**Part 2**

**Data and sources**

* + - Identify the large cities in the world with significant number of Argentinian residents.
      * This data will be sourced primarily from Wikipedia where latest city’s census can be found.
    - Identify the number of venues under the “Argentinian Restaurant” category, along with other categories that may be relevant for comparison purposes.
      * Information available in Foursquare will leverage this analysis
    - Review international tourism data visiting Buenos Aires on a yearly basis.
      * This information will be sourced by INDEC, the Argentina statistics office.

**Data analysis and statistical applications**

* + - Perform a series of correlation and regression analysis to identify any existing relationship between the number of Argentinian residents and the number of Argentinian restaurants.
    - Identify locations where a lower number of Argentinian restaurants results in a business opportunity.

**Data Sources**

* **Argentinian living abroad source data:**
  + [**https://es.wikipedia.org/wiki/Emigraci%C3%B3n\_argentina**](https://es.wikipedia.org/wiki/Emigraci%C3%B3n_argentina)
* **Argentina international tourism statistics :** 
  + [**https://www.indec.gob.ar/uploads/informesdeprensa/eti\_02\_19.pdf**](https://www.indec.gob.ar/uploads/informesdeprensa/eti_02_19.pdf)
  + **https://tradingeconomics.com/argentina/tourist-arrivals**
* **Mexico international tourism statistics:** 
  + [**https://www.yvera.tur.ar/estadistica/documentos/descarga/5a0c8e8d4c5c0.pdf**](https://www.yvera.tur.ar/estadistica/documentos/descarga/5a0c8e8d4c5c0.pdf)
* [**Argentina and New York tourisim partnership**](https://business.nycgo.com/press-and-media/press-releases/articles/post/new-york-city-and-buenos-aires-sign-first-ever-city-to-city-tourism-partnership/)
  + [**https://business.nycgo.com/press-and-media/press-releases/articles/post/new-york-city-and-buenos-aires-sign-first-ever-city-to-city-tourism-partnership/**](https://business.nycgo.com/press-and-media/press-releases/articles/post/new-york-city-and-buenos-aires-sign-first-ever-city-to-city-tourism-partnership/)
* **Other relevant data**
  + [**https://business.nycgo.com/press-and-media/press-releases/articles/post/new-york-city-and-buenos-aires-sign-first-ever-city-to-city-tourism-partnership/**](https://business.nycgo.com/press-and-media/press-releases/articles/post/new-york-city-and-buenos-aires-sign-first-ever-city-to-city-tourism-partnership/)
  + [**https://www.cronista.com/clase/break/Cuanto-y-en-que-gastan-los-turistas-que-llegan-a-la-Ciudad-de-Buenos-Aires--20190905-0004.html**](https://www.cronista.com/clase/break/Cuanto-y-en-que-gastan-los-turistas-que-llegan-a-la-Ciudad-de-Buenos-Aires--20190905-0004.html)
  + [**https://newsroom.mastercard.com/wp-content/uploads/2019/09/GDCI-Global-Report-FINAL-1.pdf**](https://newsroom.mastercard.com/wp-content/uploads/2019/09/GDCI-Global-Report-FINAL-1.pdf)
  + **https://tradingeconomics.com/argentina/tourist-arrivals**

**APPLIED DATA SCIENCE CAPSTONE**

**Final assessment: Tenedor Gaucho**

**By Florencia Paradeda**

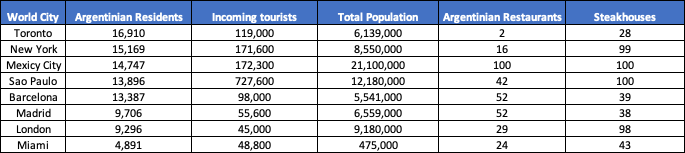
**Part 3 - Methodology Section**

**Data gathering**

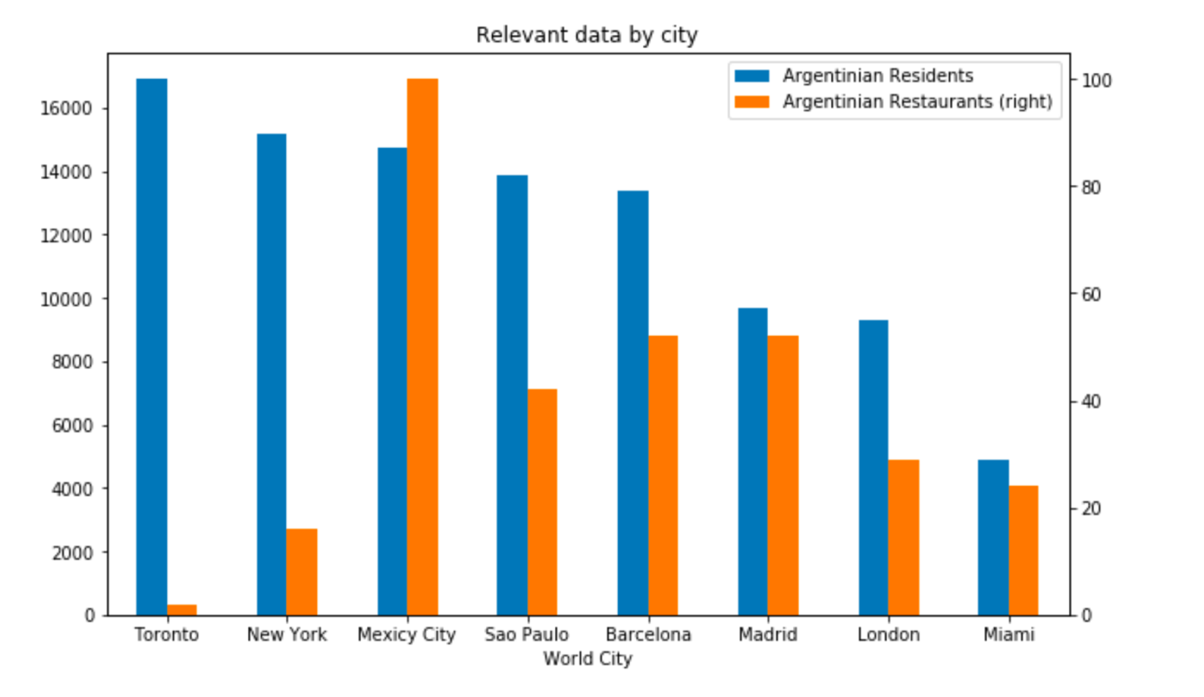
* The first step of this analysis is to obtain the relevant data for the cities that were selected for analysis.
  + Coordinates of latitude and longitude
  + Foursquare data of venues categorized as Argentinean restaurants based on geocoordinates
  + Total Population
  + Total Argentinean resident population
  + Tourists visiting Buenos Aires

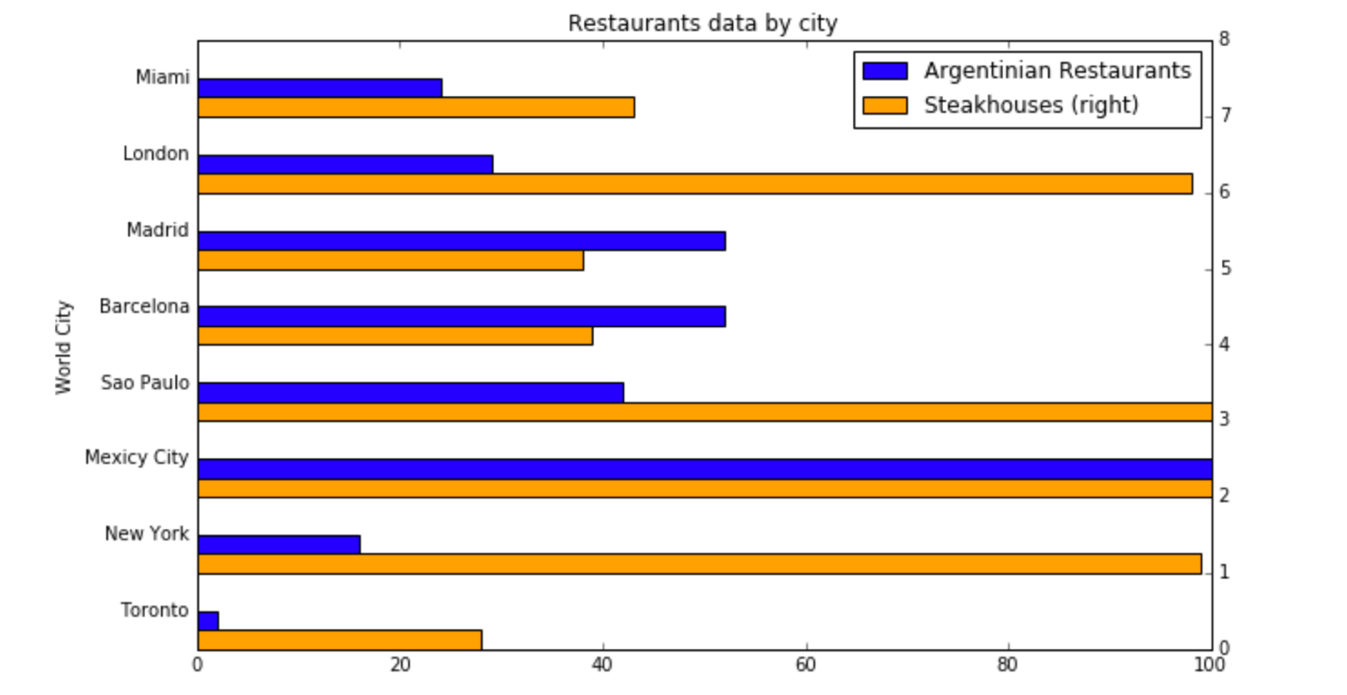
**Exploratory Data Analysis**

* After gathering the data relevant to the cities under study, I created a dataframe to consolidate the information relevant to each city.

****

* The first observation is the low number of Argentinean restaurants in both Toronto and New York compared to the number of Steakhouses and Argentinean residents. Also, the high popularity of Argentinean restaurants in Mexico City. This can be easily visualized in the following chart.





* In this first chart we can observe how the Toronto and New York markets have a much lower number of Argentinean restaurants, even though both cities are known for their active nightlife and diverse restaurant offer.
* In addition, I obtained the number of Steakhouses in each of these cities to run a comparison with the number of Argentinean restaurants. This data also shows how Toronto and New York have lower presence of Argentinean venues.

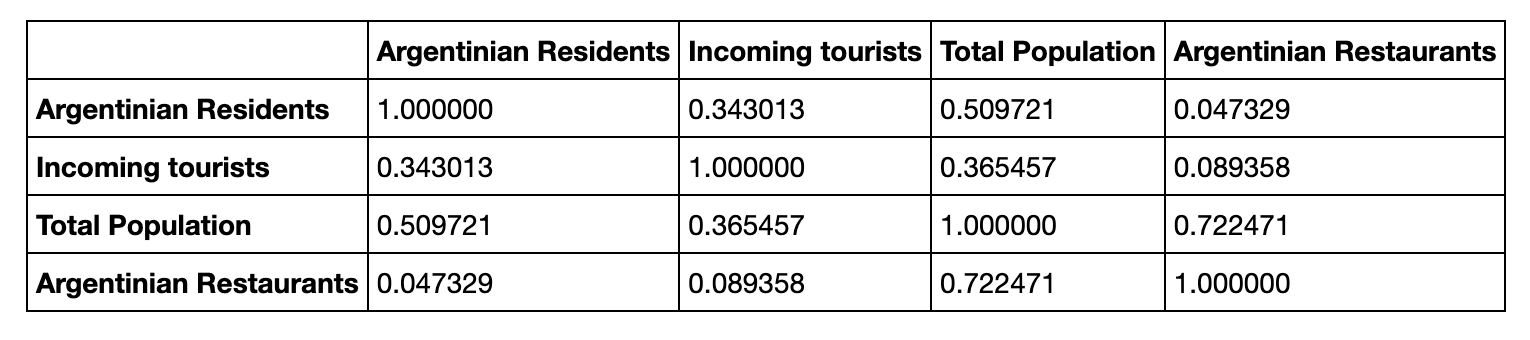


* This first observation sets our bases for the following statistical analysis.

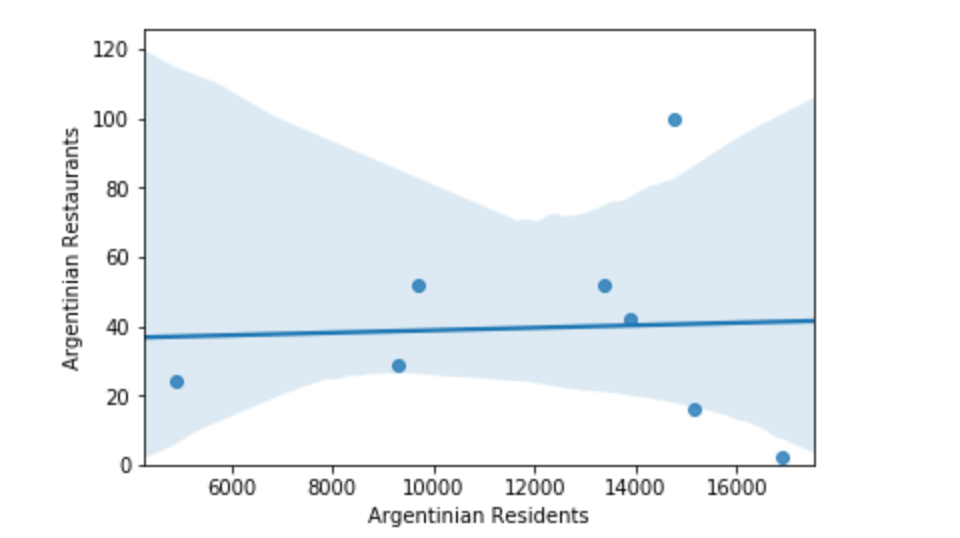
**Statistical Analysis: methodology and results**

***Considering all 8 cities***

* I will start the analysis by getting the correlation between the different data points considering all 8 cities under analysis.

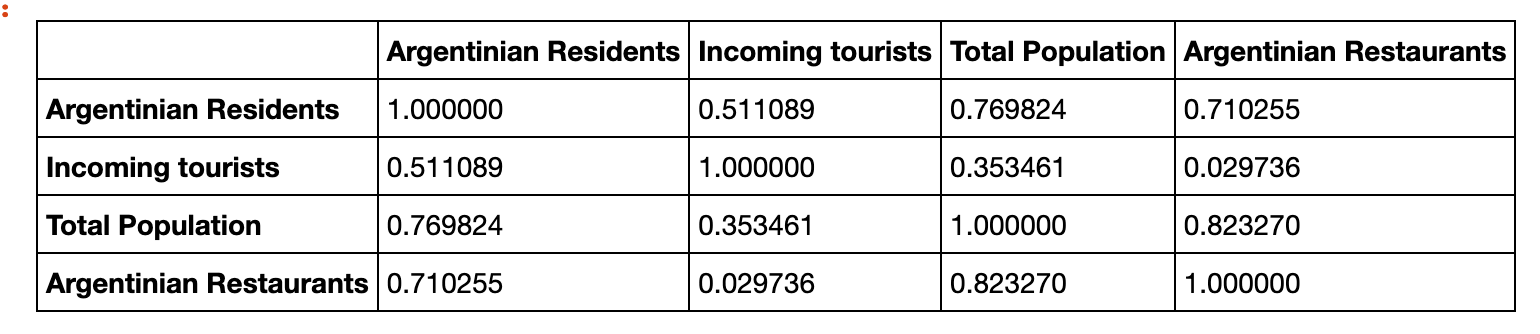


* We can see that in this analysis there is no data point that suggests a correlation of interest for further analysis, except for the total population in a particular city, which is expected. In particular, we can observe that the correlation between Argentinean residents and restaurants is about 0.05.
* Further The Pearson Correlation for 8 Cities Coefficient is 0.04732899955472767 with a P-value of P = 0.9113905599999204. This shows there is no statistical relevance.
* A visualization of the correlation between Argentinean Residents and Restaurants shows the same conclusion:

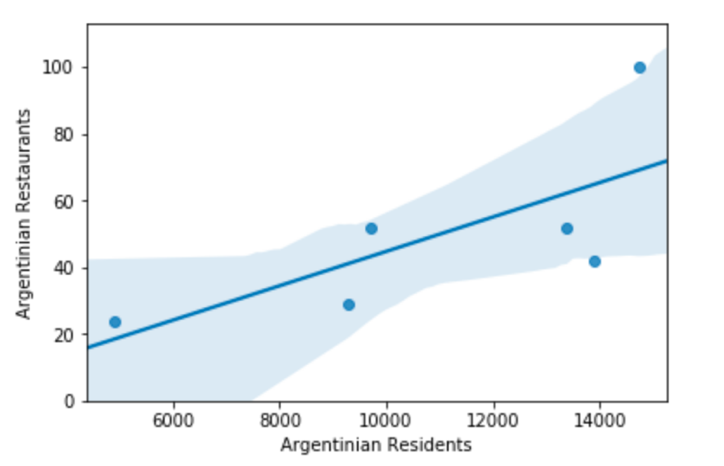


***Considering 6 cities only***

* Following the observation highlighted in the Exploratory section, I want to perform the same correlation analysis but excluding New York and Toronto.



* In this case we can observe that the correlation between Argentinean residents and restaurants in this case is over 0.7 (versus 0.05 when all 8 cities were included)
* Also, The Pearson Correlation for 6 Cities Coefficient is 0.7102554536600942 with a P-value of P = 0.11376555030152359, which represents major positive change in the statistical significance.
* We can also visualize the correlation and do a comparison with the previous analysis



**APPLIED DATA SCIENCE CAPSTONE**

**Final assessment: Tenedor Gaucho**

**By Florencia Paradeda**

**Part 4 – Results discussion, and conclusion.**

**Discussion of results**

* The basis of this analysis is the comparison of the same data for two different groups of cities, where the second group is a subset of the first group. This analysis allows for the identification of opportunities in the cities that were removed in the second group.
* We observed that cities like: Mexico City, Sao Paulo in Brazil, and Barcelona and Madrid have a very significant offer of Argentinean food. However, Toronto and New York have a much smaller offer even though both cities have a very large presence of Argentinean residents at present.
* In the analysis section we determined the correlation between residents of Argentinean origin and open restaurants serving Argentinean food for both groups of cities.
  + In the first group, where all 8 cities were included we obtained the following data:
    - Correlation: 0.05
    - Pearson Correlation for Coefficient: 0.04732899955472767
    - P-value of P = 0.911
    - This data does not support a statistical relevance between the number of Argentinean residents and restaurants.
  + However, if we look at the second group, where only 6 cities are included we obtain the following results:
    - The correlation between Argentinean residents and restaurants in this case is over 0.7. There is a much higher correlations.
    - Pearson Correlation Coefficient is 0.7102554536600942
    - P-value of P = 0.11376555030152359
* The increase in the statistical results supports the opportunity for an expansion in the offer of Argentinian restaurants in both New York and Toronto.
* If look back at the observation of number of Steakhouses in these two cities, we notice that the market in Toronto for this type of food is smaller. Therefore, New York results the location of most interest.

**Conclusion**

* New York City is the recommended location for investing in an Argentinean Restaurant.
* The provided data and statistical comparative analysis supports this recommendation.
* The attached map shows the distribution of Argentinean Restaurants in NYC

