FRIDAY, 10TH OF JULY

CAPSTONE PROJECT

THE BATTLE OF NEIGHBORHOODS

JEAN-LUC CHASLE
COURSERA
IBM Data Science

CAPSTONE PROJECT

Table des matières

Introduction / Business Problem **1**

Data **1**

Introduction/Business Problem

Bubble tea (also known as **pearl milk tea**, **bubble milk tea**, or **boba**) is a tea-based drink invented in Taiwan during the 1980s, which is shaken with ice to create the "bubbles", a foamy layer on top of the drink; chewy tapioca balls ("pearls") are added as well. Ice-blended versions are frozen and put into a blender, resulting in a slushy consistency. There are many varieties of the drink with a wide range of flavors. The two most popular varieties are black pearl milk tea and green pearl milk tea.

Culture Impact

Bubble tea has become synonymous with Taiwan and is an important symbol of Taiwanese identity both domestically and internationally. It is such a popular drink that it is often referred to as the "national drink" in Taiwan. It then spread rapidly in Asia, including China, South Korea, Malaysia, Singapore, Vietnam and Thailand. It has also been exported to many European and American cities, especially those with large Asian communities.
In France, Bubble Tea Salons have been open for many years in Paris and also in Lyon, Lille, Grenoble.

There are currently always a few number of Bubble tea shop in Paris and we would like to recommand someone who is looking to open a Bubble tea shop in french cities particularly in Paris in basis of the neighborhood (20 districts). So where will we recommend the best position that they open it?

Data

For our project, we will use the geographical position of each district in Paris in order to see if there are bubble tea shop in a district with the Foursquare API (venues), if it is a good idea to open a shop in the vicinity (if the shops around the location of the district given have good ratings, etc..) or if there are none shops.

First of all we retrieve a dataset of the geographical locations of the 20 districts of Paris, with the following features: number of the district, Latitude Venue, Longitude Venue.

Then with these elements, we use the Foursquare API to search the Bubble tea venues around each district. We get the name, categories, latitude and longitude of each venue. (We can also retrieve other features in the json file.

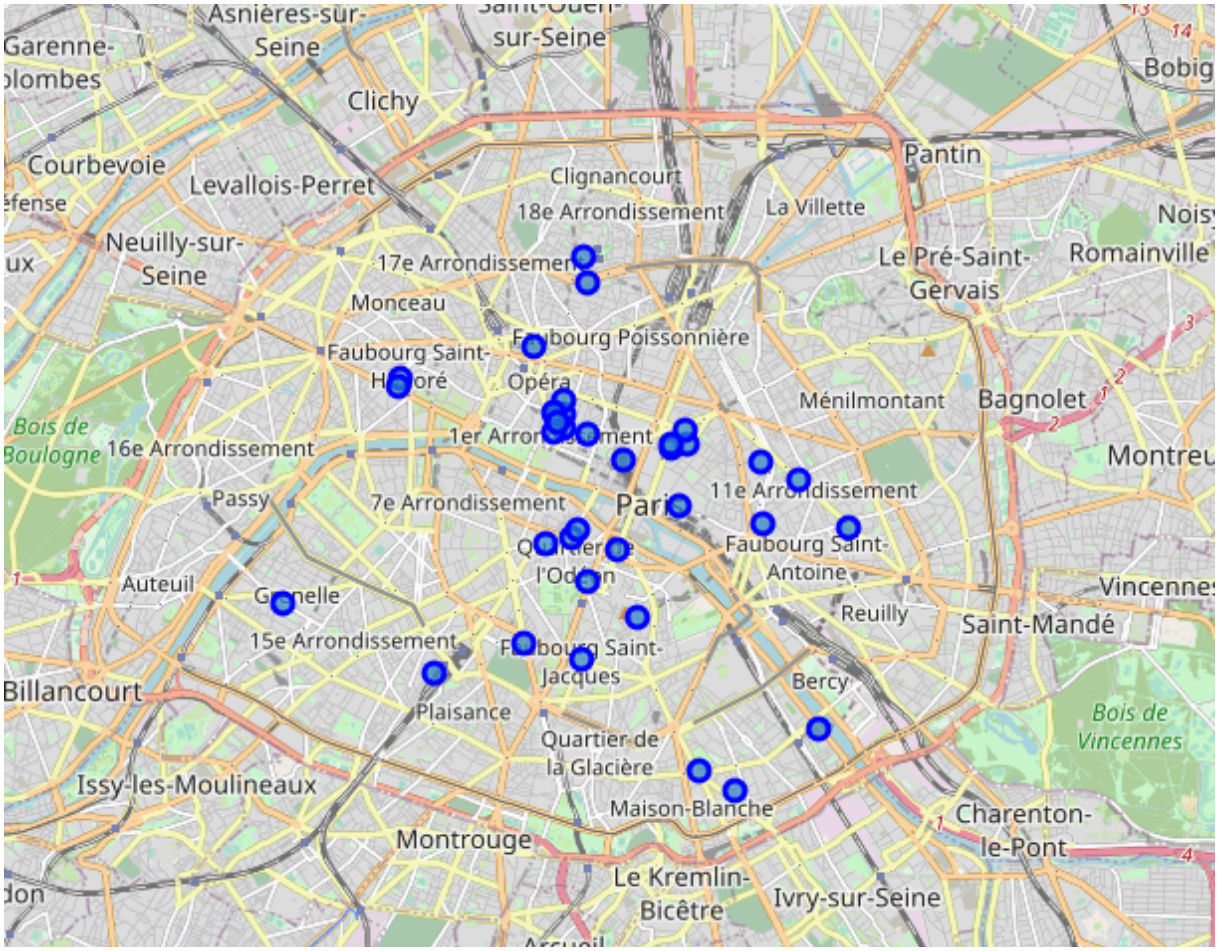
Methodology section

We calculate the number of venues per district in order to have an idea of the number of bubble tea shop per district (frequency) and then we affect a rate of 1 if there are more than 3 bubble tea in a district and 0 otherwise.
We can apply a classification model in other to predict if we give to the model a random shop with features (random geographical location of a street that is it a good idea to open the shop in that location).

Results section + Conclusion

With that results, we can deduce that in the 2, 7, 9 and 16th district, there are none bubble tea shops. So we can advice a person to open a store in their location because of the none competition in these district.

Thanks to the classification model, we deduce that we can open a bubble tea shop in the 7th district, particularly around the Invalides because the model predict that the vicinity of the district have none bubble tea shops. It is logic thanks to the visualization of the datas in the paris map.



KNN

Entrée [356]:
```
from sklearn.neighbors import KNeighborsClassifier
k = 3
#Train Model and Predict
neigh = KNeighborsClassifier(n_neighbors = k).fit(X_train,Y_train)
neigh
```

Out[356]:
```
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
           metric_params=None, n_jobs=None, n_neighbors=3, p=2,
           weights='uniform')
```

Entrée [358]:
```
X_test = [[-0.60697698, 0.47470274, 0.09059069, -0.23267278]]
yhat = neigh.predict(X_test)
print(yhat)
```

```
[0.]
```

Entrée [359]:
```
# With these randoms values of the location, we can deduce that it will be less than 3 bubble tea shops at that loca
```

Entrée [376]:
```
# Geographical location of the invalides in the 7th district
X_test_district_7th = pd.DataFrame({'Arrondissement':[7.0],'Venue':[0.0], 'Latitude': [48.8584991851266], 'Longitude
X_test2 = preprocessing.StandardScaler().fit(X_test_district_7th).transform(X_test_district_7th)
X_test2
```

Out[376]:
```
array([[0., 0., 0., 0.]])
```

Entrée [377]:
```
yhat2 = neigh.predict(X_test2)
print(yhat)
```

```
[0.]
```

Entrée [378]:
```
# It is a good idea to open a bubble tea shop around the invalides
```

Entrée [ ]: