# Statement

## Part 1: Description of the problem

Let's imagine that a certain person, let's call him Yuddy. Yuddy has his habits and he enjoys going to some venues. For example, he can enjoy going to the cinema, or eating in a Vietnamese restaurant. The places where you go can sometimes describe the person you are. Yuddy really enjoys going to those places but sometimes he could feel tired of doing always the same plans.

So now Yuddy is tired of going to the same Vietnamese restaurant and he wants to get to know other places where he can eat. But not other random venues, but venues that could fit with his tastes.

The problem that we will discuss here is knowing which of the venues in his city could fit Yuddy's likes. This is a particular case that, however, could be extrapolated to any other client that asked us for help.

So, in short, we could have some kind of business which could help **any person interested in getting to know new venues that could fit their likes**.

## Part 2: Data used to solve the problem

So how do we find this venues? I think a **Collaborative Filtering recommendation engine** could be perfect to solve this problem. This engine takes other users and find the similarity between their likes and Yuddy's likes. With those similaritys and differences, the engine is able to recommend new places unvisited by Yuddy.

But how do we choose those other users in order to compare their likes with Yuddy's ones? Well, there could be several ways. The method that we will be using is finding a venues which Yuddy liked. In this venue we are going to pick **up to 10 users** that has commented a tip. This 10 users are those that will be used to fit the collaborative filtering recommendation engine.

Now that we have the users selected, what we are going to do is look which other places have these users visited, so we could create our collaborative filtering recommendation engine.

# Development

## Finding Yuddy's likes

So first thing we do is retrieving information about Yuddy with the Foursquare API. With it, we find which places has Yuddy visited and leaved a tip:

| | text | venue.name | venue.categories |
|---|---|---|---|
| 0 | Playground for children. | Grand Park | [{'id': '4bf58dd8d48988d163941735', 'name': 'Park', 'pluralName': 'Parks', 'shortName': 'Park', 'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/parks_outdoors/park_', 'suffix': '.png'}, 'primary': True}] |
| 1 | Home made ice cream without preservatives.. | Ragusa Es Italia Restaurant & Ice Cream | [{'id': '4bf58dd8d48988d1c9941735', 'name': 'Ice Cream Shop', 'pluralName': 'Ice Cream Shops', 'shortName': 'Ice Cream', 'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/icecream_', 'suffix': '.png'}, 'primary': True}] |
| 2 | It is within MKG shopping mall where a place of many great restaurant with traditional and international cuisines. | HARRIS Hotel & Conventions Kelapa Gading | [{'id': '4bf58dd8d48988d1fa931735', 'name': 'Hotel', 'pluralName': 'Hotels', 'shortName': 'Hotel', 'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/travel/hotel_', 'suffix': '.png'}, 'primary': True}] |
| 3 | It is within MKG Shopping Mall next to several | HARRIS Hotel & Conventions Kelapa Gading | [{'id': '4bf58dd8d48988d1fa931735', 'name': 'Hotel', 'pluralName': 'Hotels', 'shortName': 'Hotel', 'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/travel/hotel_', 'suffix': '.png'}, 'primary': True}] |
| 4 | Free parking, near shopping mall, convenient.. | HARRIS Hotel & Conventions Kelapa Gading | [{'id': '4bf58dd8d48988d1fa931735', 'name': 'Hotel', 'pluralName': 'Hotels', 'shortName': 'Hotel', 'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/travel/hotel_', 'suffix': '.png'}, 'primary': True}] |
| 5 | Fast wifi, authentic affordable food, complimentary massage, but not good for overnite rest since there is no heater and not many money changers available. | Taiwan Taoyuan International Airport (TPE) (臺灣桃園國際機場) | [{'id': '4bf58dd8d48988d1ed931735', 'name': 'Airport', 'pluralName': 'Airports', 'shortName': 'Airport', 'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/travel/airport_', 'suffix': '.png'}, 'primary': True}] |

Table 1: Yuddy's tips

We should clean the table in order to be able to work with its values. So now we drop the unnecessary information in the venue.categories column in order to obtain just the categorie of the venue visited by Yuddy (that is what we want to find). Let's also rename the column to some more intuitive titles and take a look to the first 9 results:

| | Text | Name | Categorie |
|---|---|---|---|
| 0 | Playground for children. | Grand Park | Park |
| 1 | Home made ice cream without preservatives.. | Ragusa Es Italia Restaurant & Ice Cream | Ice Cream Shop |
| 2 | It is within MKG shopping mall where a place of many great restaurant with traditional and international cuisines. | HARRIS Hotel & Conventions Kelapa Gading | Hotel |
| 3 | It is within MKG Shopping Mall next to several | HARRIS Hotel & Conventions Kelapa Gading | Hotel |
| 4 | Free parking, near shopping mall, convenient.. | HARRIS Hotel & Conventions Kelapa Gading | Hotel |
| 5 | Fast wifi, authentic affordable food, complimentary massage, but not good for overnite rest since there is no heater and not many money changers available. | Taiwan Taoyuan International Airport (TPE) (臺灣桃園國際機場) | Airport |
| 6 | Cozy place, nice stores, free wifi, but not a good resting area. The area has no heater during the nite. | Terminal 2 (臺灣桃園國際機場第二航廈) | Airport Terminal |
| 7 | One of the top 5 liberal arts colleges in the US. | Pomona College | University |
| 8 | Phily steak with chicken soup. It comes with 2 subs. | Euro Cafe | Restaurant |
| 9 | They usually have 2-day flash sale. Sometime up to 50%. | Huntley Bookstore | College Bookstore |

Table 2: Yuddy's tips cleaned

So now we can see the places that Yuddy has visited. We could have now some idea of Yuddy's behaviour. We could imagine he is an academic that enjoys traveling and eating outside (both in restaurants and ice creams).

As we see, it seems that he could spend more time in Hotels than in Parks. So maybe we could deduct that, although he likes both places, he may prefer hotels to parks. In order to get to know Yuddy better, we are going to construct a new table with the categories of the venues that Yuddy has visited and the frequency that he has visited each one:

| | Categorie | Favorite |
|---|---|---|
| 0 | Park | 1 |
| 1 | Ice Cream Shop | 2 |
| 2 | Hotel | 3 |
| 3 | Airport | 1 |
| 4 | Airport Terminal | 1 |
| 5 | University | 1 |
| 6 | Restaurant | 2 |
| 7 | College Bookstore | 2 |
| 8 | Bakery | 1 |
| 9 | Frozen Yogurt Shop | 2 |
| 10 | American Restaurant | 1 |
| 11 | Coffee Shop | 2 |
| 12 | Mexican Restaurant | 2 |
| 13 | Gastropub | 1 |
| 14 | Vietnamese Restaurant | 2 |
| 15 | Thai Restaurant | 1 |
| 16 | Indian Restaurant | 1 |
| 17 | Asian Restaurant | 1 |
| 18 | Breakfast Spot | 1 |
| 19 | Donut Shop | 1 |
| 20 | Supermarket | 1 |

Table 3: Yuddy's favorite categories

Now we can construct a more precise idea of Yuddy. We see that, indeed, he likes to travel (based on the airport and hotel results). And it seems that he also enjoys eating in several types of venues (from Donut Shops to Mexican restaurant).

Imagine that we asked Yuddy to punctuate each venue category from 0 to 5, where 5 would be his favorite category and 0 would be a category that he didn´t enjoyed that much. Based on the frequency of his visits, maybe he would tell us that he prefers the Vietnamese or Mexican restaurants that the Thai ones. So we could make an approximation to Yuddy's likes with the information of the frequency of his visits where 5 would be the most visited place. The rating of the other places would be five times its frequency divided by the frequency of the most visited place. With this reasoning we are able to construct this new table:

| | Categorie | Favorite |
|---|---|---|
| 0 | Park | 1.666667 |
| 1 | Ice Cream Shop | 3.333333 |
| 2 | Hotel | 5.000000 |
| 3 | Airport | 1.666667 |
| 4 | Airport Terminal | 1.666667 |
| 5 | University | 1.666667 |
| 6 | Restaurant | 3.333333 |
| 7 | College Bookstore | 3.333333 |
| 8 | Bakery | 1.666667 |
| 9 | Frozen Yogurt Shop | 3.333333 |
| 10 | American Restaurant | 1.666667 |
| 11 | Coffee Shop | 3.333333 |
| 12 | Mexican Restaurant | 3.333333 |
| 13 | Gastropub | 1.666667 |
| 14 | Vietnamese Restaurant | 3.333333 |
| 15 | Thai Restaurant | 1.666667 |
| 16 | Indian Restaurant | 1.666667 |
| 17 | Asian Restaurant | 1.666667 |
| 18 | Breakfast Spot | 1.666667 |
| 19 | Donut Shop | 1.666667 |
| 20 | Supermarket | 1.666667 |

Table 4: Yuddy's favorite categories with ratings

Now we already have Yuddy's tastes and we can look for other users with similar ones.

## Finding others likes

Fist thing we have to do is finding other users with similar tastes. In order to do so we pick one of Yuddy favorite places (Name column in Table 2) and retrieve other users tips. We pick up to 10 users that has visited that place. This users could share some preferences with Yuddy.

As we did with with the process form Table 1 to Table 3, we clean and discriminate the data so we are left only with the relevant information. Let's take a look to the first 5 rows of this research:

| | User Name | Categorie |
|---|---|---|
| 0 | GloriaA | Observatory |
| 1 | GloriaA | Korean Restaurant |
| 2 | GloriaA | Park |
| 3 | Brandon Featherstone | Breakfast Spot |
| 4 | Brandon Featherstone | Mexican Restaurant |

Table 5: Users venues

We can also examine which kind of places have visited each one of the users. Now we look for example to the places visited by Kathia Concepcion:

| | User Name | Categorie |
|---|---|---|
| 50 | Kathia Concepcion | Art Museum |
| 51 | Kathia Concepcion | Nightclub |
| 52 | Kathia Concepcion | Park |
| 53 | Kathia Concepcion | Park |
| 54 | Kathia Concepcion | Market |
| 55 | Kathia Concepcion | Salon / Barbershop |
| 56 | Kathia Concepcion | Hotel |
| 57 | Kathia Concepcion | Hotel |
| 58 | Kathia Concepcion | Salon / Barbershop |
| 59 | Kathia Concepcion | Multiplex |
| 60 | Kathia Concepcion | Hotel |
| 61 | Kathia Concepcion | Mexican Restaurant |
| 62 | Kathia Concepcion | Grocery Store |
| 63 | Kathia Concepcion | Sushi Restaurant |
| 64 | Kathia Concepcion | Jazz Club |
| 65 | Kathia Concepcion | Pizza Place |
| 66 | Kathia Concepcion | Japanese Restaurant |

Table 6: Kathias favorite categories

What we are going to do now is adding a new column with the times each user has visited each venues category and we obtain this table:

| | User | Categorie | Favorite |
|---|---|---|---|
| **0** | GloriaA | Observatory | 1 |
| **1** | GloriaA | Korean Restaurant | 1 |
| **2** | GloriaA | Park | 1 |
| **3** | Brandon Featherstone | Breakfast Spot | 1 |
| **4** | Brandon Featherstone | Mexican Restaurant | 3 |
| **...** | ... | ... | ... |
| **140** | Whitney M | Trail | 1 |
| **141** | Whitney M | Café | 1 |
| **142** | Whitney M | Coffee Shop | 2 |
| **143** | Whitney M | Breakfast Spot | 1 |
| **144** | Whitney M | BBQ Joint | 1 |

Table 7: Users favorite categories

Let's check the correctness of this table by examining again Kathis slice:

| | User | Categorie | Favorite |
|---|---|---|---|
| **38** | Kathia Concepcion | Art Museum | 1 |
| **39** | Kathia Concepcion | Nightclub | 1 |
| **40** | Kathia Concepcion | Park | 2 |
| **41** | Kathia Concepcion | Market | 1 |
| **42** | Kathia Concepcion | Salon / Barbershop | 2 |
| **43** | Kathia Concepcion | Hotel | 3 |
| **44** | Kathia Concepcion | Multiplex | 1 |
| **45** | Kathia Concepcion | Mexican Restaurant | 1 |
| **46** | Kathia Concepcion | Grocery Store | 1 |
| **47** | Kathia Concepcion | Sushi Restaurant | 1 |
| **48** | Kathia Concepcion | Jazz Club | 1 |
| **49** | Kathia Concepcion | Pizza Place | 1 |
| **50** | Kathia Concepcion | Japanese Restaurant | 1 |

Table 8: Kathias favorite categories with frequency

Now we convert the Favorite column into ratings as we did with Yuddy's one:

| | User | Categorie | Favorite |
|---|---|---|---|
| 38 | Kathia Concepcion | Art Museum | 1.66667 |
| 39 | Kathia Concepcion | Nightclub | 1.66667 |
| 40 | Kathia Concepcion | Park | 3.33333 |
| 41 | Kathia Concepcion | Market | 1.66667 |
| 42 | Kathia Concepcion | Salon / Barbershop | 3.33333 |
| 43 | Kathia Concepcion | Hotel | 5 |
| 44 | Kathia Concepcion | Multiplex | 1.66667 |
| 45 | Kathia Concepcion | Mexican Restaurant | 1.66667 |
| 46 | Kathia Concepcion | Grocery Store | 1.66667 |
| 47 | Kathia Concepcion | Sushi Restaurant | 1.66667 |
| 48 | Kathia Concepcion | Jazz Club | 1.66667 |
| 49 | Kathia Concepcion | Pizza Place | 1.66667 |
| 50 | Kathia Concepcion | Japanese Restaurant | 1.66667 |

Table 9: Kathias favorite categories with ratings

## Finding common preferences

What we do know is taking a slice of Table 9 so we are picking just those categories in common with Yuddy's ones. These are the first columns:

| | User | Categorie | Favorite |
|---|---|---|---|
| 2 | GloriaA | Park | 5 |
| 3 | Brandon Featherstone | Breakfast Spot | 1.25 |
| 4 | Brandon Featherstone | Mexican Restaurant | 3.75 |
| 11 | Brandon Featherstone | Donut Shop | 2.5 |
| 13 | Brandon Featherstone | Park | 3.75 |
| 18 | Brandon Featherstone | Gastropub | 1.25 |
| 19 | Rebecca Ronquillo | Airport | 2.5 |
| 22 | Rebecca Ronquillo | Indian Restaurant | 2.5 |
| 23 | Rebecca Ronquillo | Mexican Restaurant | 2.5 |
| 25 | Rebecca Ronquillo | Park | 2.5 |
| 31 | Rebecca Ronquillo | Coffee Shop | 5 |
| 36 | Gladis Diaz | Park | 5 |
| 40 | Kathia Concepcion | Park | 3.33333 |
| 43 | Kathia Concepcion | Hotel | 5 |
| 45 | Kathia Concepcion | Mexican Restaurant | 1.66667 |
| 53 | Noel Holmes | Park | 1.66667 |
| 58 | Noel Holmes | Coffee Shop | 3.33333 |
| 59 | Noel Holmes | Hotel | 1.66667 |
| 61 | Patrick | Bakery | 1.25 |
| 63 | Patrick | Hotel | 1.25 |
| 67 | Patrick | Airport | 3.75 |

Table 10: Common categories with ratings

Next we group the table by users in order to get the top 3 users that share the most common venues with Ana. Doing so we obtain that these users are:

- Whitney M: with 8 venues
- Patrick: with 7 venues
- Brandon Featherstone: with 5 venues

Now we are going to find which user is most similar to Yuddy by using the Pearson Correlation Coefficient, used to measure the strength of the relation between two linear variables.

This coefficient will return a value between -1 and 1 where -1 would be a perfect negative correlation (the oposite) and 1 would mean that both users have similar tastes.

By applying this coefficient to the data we obtain Table 11 with the following heading:

| | similarityIndex |
|---|---|
| Rebecca Ronquillo | 0.612372 |
| Brandon Featherstone | 0.559017 |
| Kathia Concepcion | 0.500000 |
| Christopher Jones | 0.218218 |
| Noel Holmes | 0.000000 |

Table 11: Users and their Similarity Index

## Getting recommendations

Now that we have a table with the users and their similarity index with Yuddy we can construct a new table where a column indicates the weighted rating for each row. This weighted rating is obtained by multiplying the similarity index by the users ratings. Doing so we obtain a table as follows:

| | similarityIndex | User | Categorie | Favorite | weightedRating |
|---|---|---|---|---|---|
| 0 | 0.612372 | Rebecca Ronquillo | Airport | 2.5 | 1.53093 |
| 1 | 0.612372 | Rebecca Ronquillo | Salon / Barbershop | 2.5 | 1.53093 |
| 2 | 0.612372 | Rebecca Ronquillo | Gas Station | 2.5 | 1.53093 |
| 3 | 0.612372 | Rebecca Ronquillo | Indian Restaurant | 2.5 | 1.53093 |
| 4 | 0.612372 | Rebecca Ronquillo | Mexican Restaurant | 2.5 | 1.53093 |

Table 12: Users and their weighted ratings

We observe that each rows contain one user, one place visited by that user, his similarity index, the rating and the weighted rating using the similarity index.

We group now this new table by the category and sum the similarity index, so we can obtain a table in which each rows contains the information about one venue:

| Categorie | sum_similarityIndex | sum_weightedRating |
|---|---|---|
| Airport | -0.064977 | -1.167880 |
| American Restaurant | -0.332132 | -2.388051 |
| Art Gallery | 0.559017 | 0.698771 |
| Art Museum | 1.059017 | 1.532105 |
| Asian Restaurant | -0.127000 | -0.211667 |

Table 13: Categories and their weighted ratings

Now we can obtain the recommendations. To do so we just divide the weighted rating of each venue by its similarity index, obtaining the weighted average recommendation score. So finally we obtain the following sorted table:

| Categorie | weighted average recommendation score | Categorie |
|---|---|---|
| Burger Joint | 243.110351 | Burger Joint |
| Office | 38.274982 | Office |
| Airport | 17.973686 | Airport |
| Taco Place | 13.591661 | Taco Place |
| American Restaurant | 7.190074 | American Restaurant |
| New American Restaurant | 5.653790 | New American Restaurant |
| Latin American Restaurant | 5.000000 | Latin American Restaurant |
| Coffee Shop | 4.783923 | Coffee Shop |
| Park | 3.728594 | Park |
| Residential Building (Apartment / Condo) | 3.333333 | Residential Building (Apartment / Condo) |

Table 14: Categories and their recommended score

Now we can drop those places that wouldn't fit anyone's preferences (who would like to go to the office?) and those already visited by Yuddy. If we pick the top 3 recommended venues by doing so, we obtain:

- Burger Joint
- Taco Place
- Latin American Restaurant

## Examining results

So we have obtained a Burger Joint, a Taco Place and a Latin American Restaurant as the top 3 recommended venues to go (we could continue looking at the following rows of the table to find more venues). It seems these places fits with what we know so far about Yuddys preferences. We hope he'll like them!