**Report  
City for a pleasant life**

[1 A description of the problem 2](#_Toc535693509)

[1.1 Introduction 2](#_Toc535693510)

[1.2 Audience of my project and why they would care about this problem 2](#_Toc535693511)

[2. Discussion of the background 3](#_Toc535693512)

[3. A description of the data 4](#_Toc535693513)

[3.1 Data sources 4](#_Toc535693514)

[3.2 Data Features: 4](#_Toc535693515)

[4. How it will be used to solve the problem 5](#_Toc535693516)

[5. Methodology section¶ 6](#_Toc535693517)

[5.1. Download data on areas of the city of Toronto. 6](#_Toc535693518)

[5.2. Prepare 'data' - a new DataFrame to store data on venues in Toronto 6](#_Toc535693519)

[5.3. Obtain data using Foursquare API 6](#_Toc535693520)

[5.4. Get all Venue Category from data 7](#_Toc535693521)

[5.5. We will create categories of venues that each family member likes - husband (F), wife (W), daughter (D) 8](#_Toc535693522)

[5.6. For each venue we add its category (column 'Cat') to 'data' - F, B or D 8](#_Toc535693523)

[5.7. Find the number of venues that each family member likes in each of the city districts 9](#_Toc535693524)

[5.8. Add the values found to "df" common data frame 9](#_Toc535693525)

[6. Results section 10](#_Toc535693526)

[6.1. Family priority scheme (f >> w > d) (0.7, 0.2, 0.1) 10](#_Toc535693527)

[6.2. Family priority scheme (f < w << d) (0.1, 0.2, 0.7) 10](#_Toc535693528)

[6.3. Family priority scheme (w >> f > d) (0.2, 0.7, 0.1) 10](#_Toc535693529)

[6.4. Family priority scheme (f = w = d) (1/3, 1/3, 1/3) 10](#_Toc535693530)

[6.5. Imagine the data in graphical form 11](#_Toc535693531)

[6.6. Imagine the data in graphical form 11](#_Toc535693532)

[6.7. Imagine the Top-5 data in graphical form for each scheme 12](#_Toc535693533)

[7. Discussion section 13](#_Toc535693534)

[8. Conclusion section 14](#_Toc535693535)

# 1 A description of the problem

## Introduction

I have a good friend. And I want to help him.

My friend is a famous artist. And now he is looking for a city to which he could go to work. Of course, first of all he is interested in art. As for all people of art, the environment, the atmosphere in which he will live, means a lot to him.

Suppose they have the opportunity to move to one of several cities. In one of the areas of these cities. Let it be Toronto and his districts (or Florence, Stockholm or Minsk ...).

I want, using data analysis, to determine which district of which city will be the most pleasant for my friend's family.

His family consists of three people - himself, his wife and daughter. And he wants each of them to get what he likes in the new city.

His wife is a very good cook. He knows the cuisine of different countries. But he wants to know even more.

Their daughter loves all the tasty things - juices, ice cream, fruit.

My friend is a subtle connoisseur of wine and coffee. But does not like beer.

## Audience of my project and why they would care about this problem

The audience of my project is people who want to change their place of residence. Or just tourists who want to choose not just a place where you can live for several weeks.

For this, sites [https://www.booking.com](https://www.booking.com/) or [https://www.airbnb.ru](https://www.airbnb.ru/) are enough. And the place where it can be done with the greatest pleasure.

Why will they take care of this problem?

Because it is new for them, interesting and necessary. And it costs almost nothing)

# 2. Discussion of the background

It is clear that the wife agrees to live in the district of city where there are many cafes and a variety of restaurants - Italian, Indian, Chinese, Japanese ...

Daughter - where there are many children's cafes, ice cream cafes, flower shops ...

My friend - where there are liquor stores, bars, but not pubs)

\* \* \* \* \*

Obviously, in order to be able to make a decision, the task must be formalized.

Let the number of places in the area, pleasant for the wife - W. Number of places pleasant for a child - D. Number of places pleasant for a friend - F.

We will also assume that the result of the choice of each family member affects to varying degrees. The coefficient for choosing a friend is f, wives — w, daughters — d. And f + c + d = 1.

It is also important to consider how many people live in the area. If there are a lot of them, this is not the best option.

Thus, the task is reduced to the fact that you need to find a DS district in one of the cities S, such that

**R (S, DS) = (F \* f + W \* w + D \* d) / P (\*)**

where P is the population of the district, will be the greatest.

You can even more detail:

**R (S, DS1, ..., DSn) = (F \* f + W \* w + D \* d) / P (\*\*)**

# 3. A description of the data

## 3.1 Data sources

Wikipedia (city's districts, population):   
[https://en.wikipedia.org](https://en.wikipedia.org/)

Google Maps Geocoding API:   
<http://geoawesomeness.com/developers-up-in-arms-over-google-maps-api-insane-price-hike/>

Geographical coordinates for postal codes:   
<http://cocl.us/Geospatial_data>

Foursquare API:   
[https://api.foursquare.com](https://api.foursquare.com/)

## Data Features:

We will be leveraging on features in a reliable location information provider such as the Foursquare.com to explore the various types of venues and its categories available in each neighbourhood. We will also need to understand the type of these venues nearby (i.e. within 1000M) in each of the respective neighbourhood. The information obtained per neighbourhood will be as such like below and must be in a structured format so to allow for further computation:

1. Neighbourhood
2. Neighbourhood Latitude
3. Neighbourhood Longitude
4. Venue Name
5. Venue Category
6. Venue Latitude
7. Venue Longitude

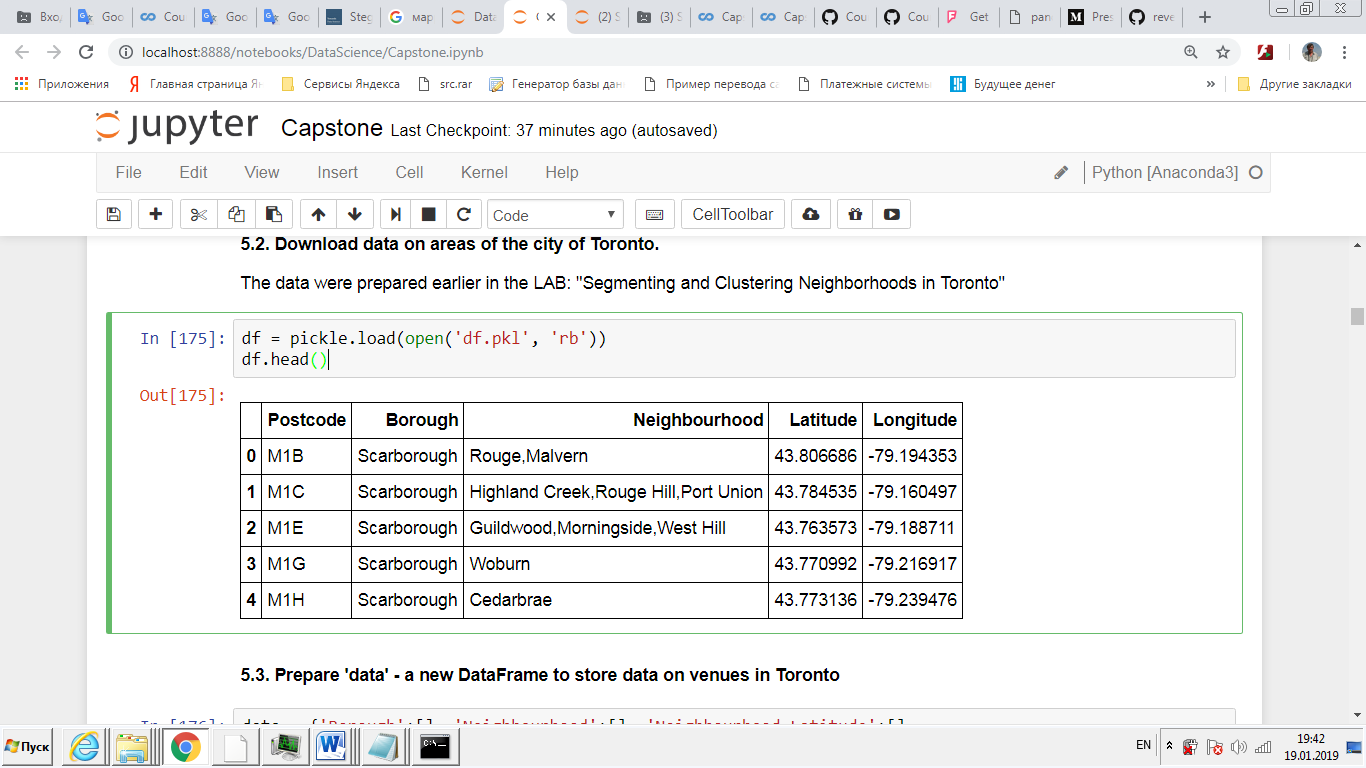
# 4. How it will be used to solve the problem

* 1. From the pages of Wikipedia to obtain data on the administrative-territorial division of cities.
  2. About the population of cities and their areas.
  3. Associate the information found with the geographical coordinates of the areas.
  4. Get, using the service Foursquare, information about places located in these areas.
  5. Select those places that interest family members, and count their number.
  6. Apply the formula (\**) or (\*\**) to the found information.
  7. Choose the best result and recommend it to my friend.
  8. Listen to his comments (or suggestions) and try to search for information in a new way.

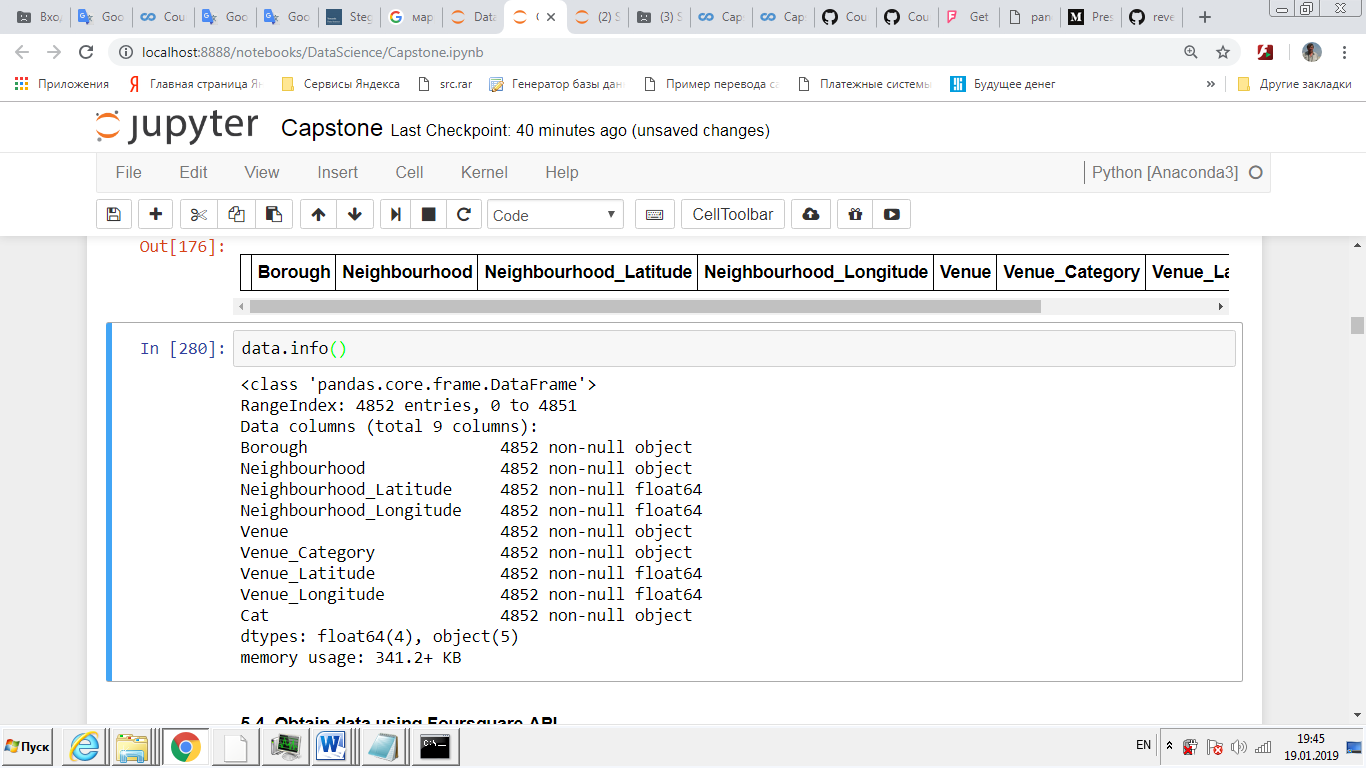
# 5. Methodology section[¶](http://localhost:8888/notebooks/DataScience/Capstone.ipynb#5.-Methodology-section)

## 5.1. Download data on areas of the city of Toronto.

The data were prepared earlier in the LAB: "Segmenting and Clustering Neighborhoods in Toronto"



## 5.2. Prepare 'data' - a new DataFrame to store data on venues in Toronto



## 5.3. Obtain data using Foursquare API

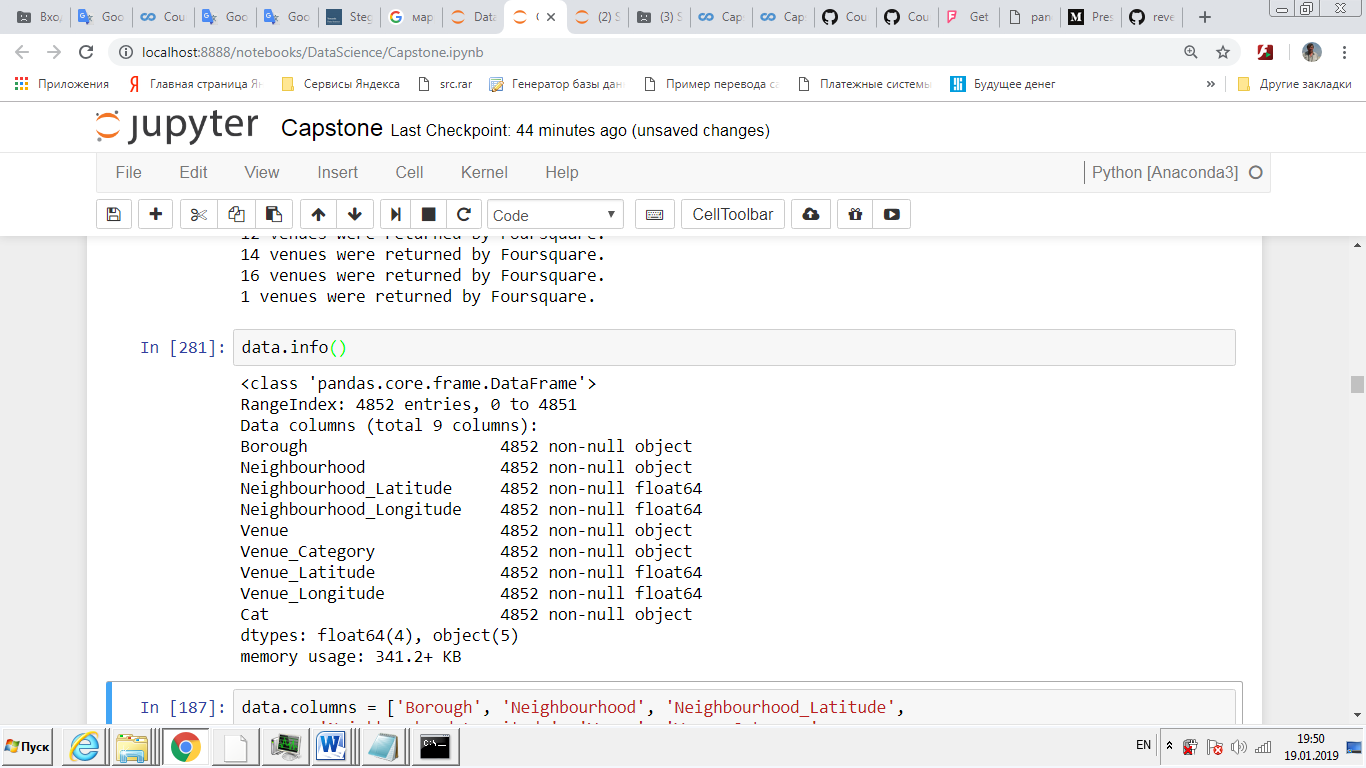
Data scrapping from the Wikipedia page that contains informaion of Toronto neighbourhoods has been used. This is critical to understand of each Toronto neighbourhood which is one of the key elements in the neighbourhood of choice in this project.

We need to know the coordinates and locations of this neighbourhoods, and therefore the geolocator API has been used for achieving this objective. This is important so that we can input this information into the location information provider such as Foursquare.com to obtain venue information in these neighbourhoods.

Finally, with all these methodologies, we will then be able to come up with a best recommendation to family of my friend to their problem.

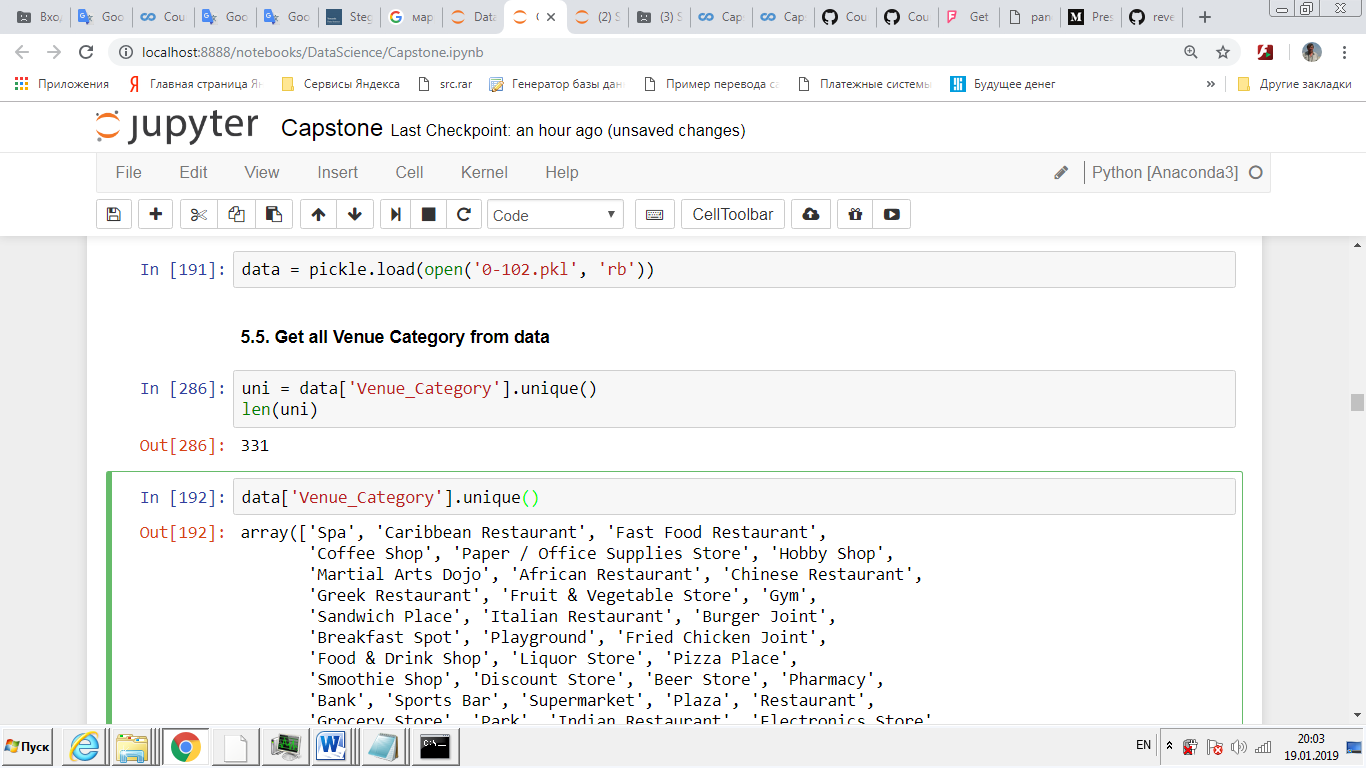
LIMIT = 1000 # limit of number of venues returned by Foursquare API

Radius = 1000 # define radius

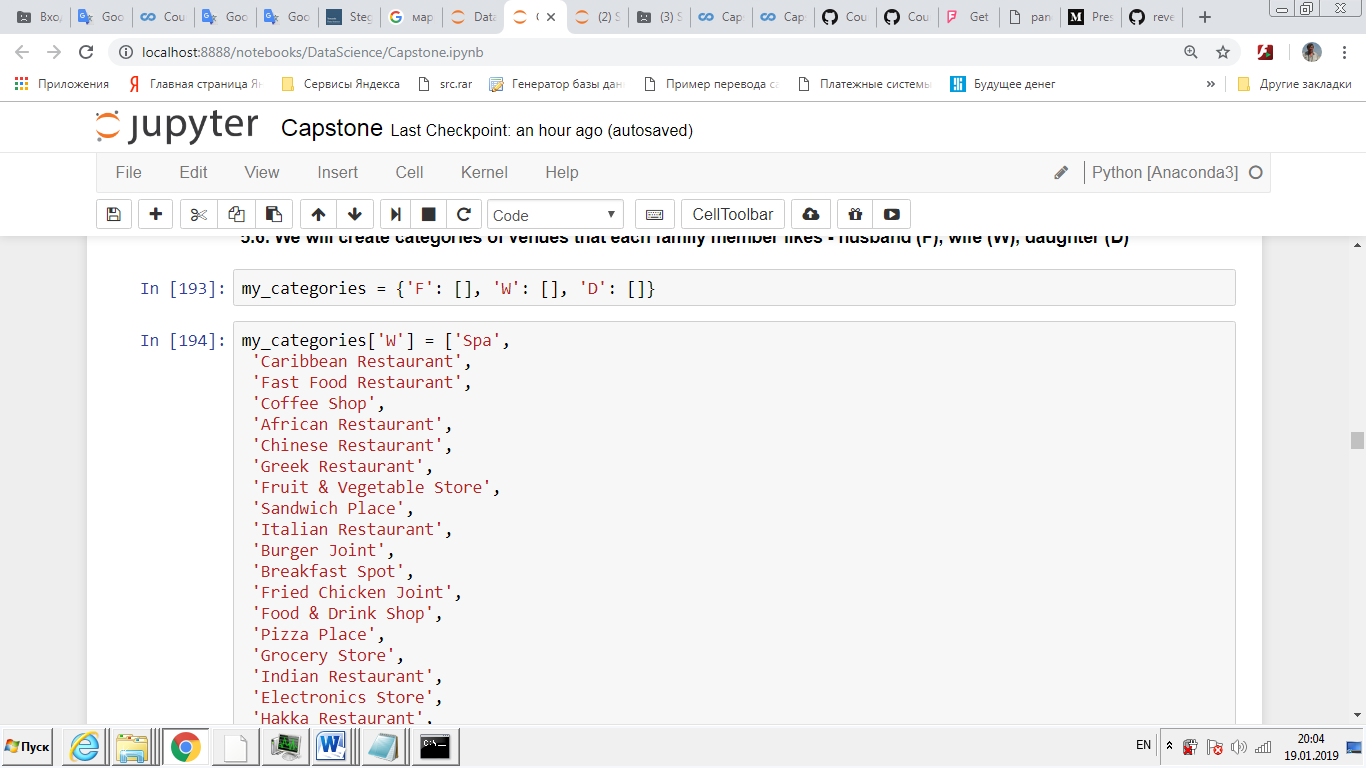


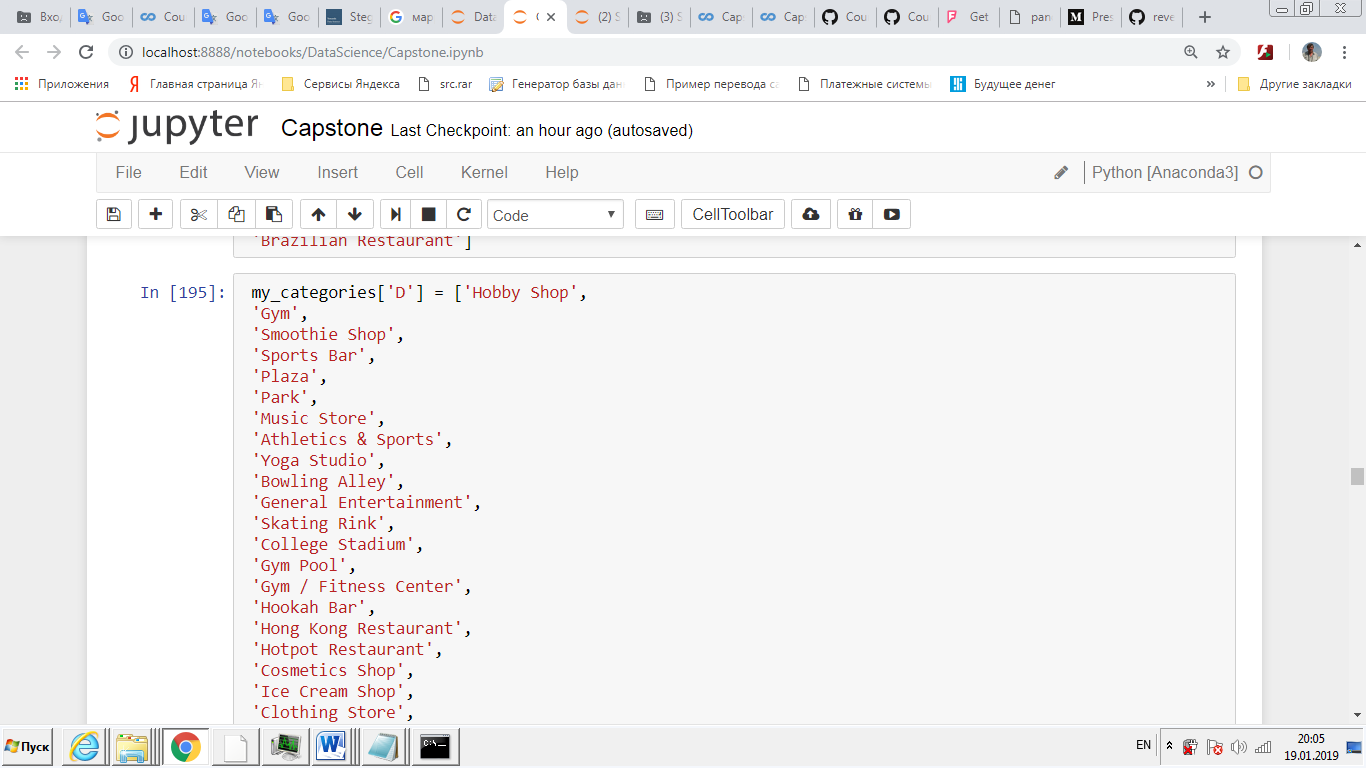
## 5.4. Get all Venue Category from data

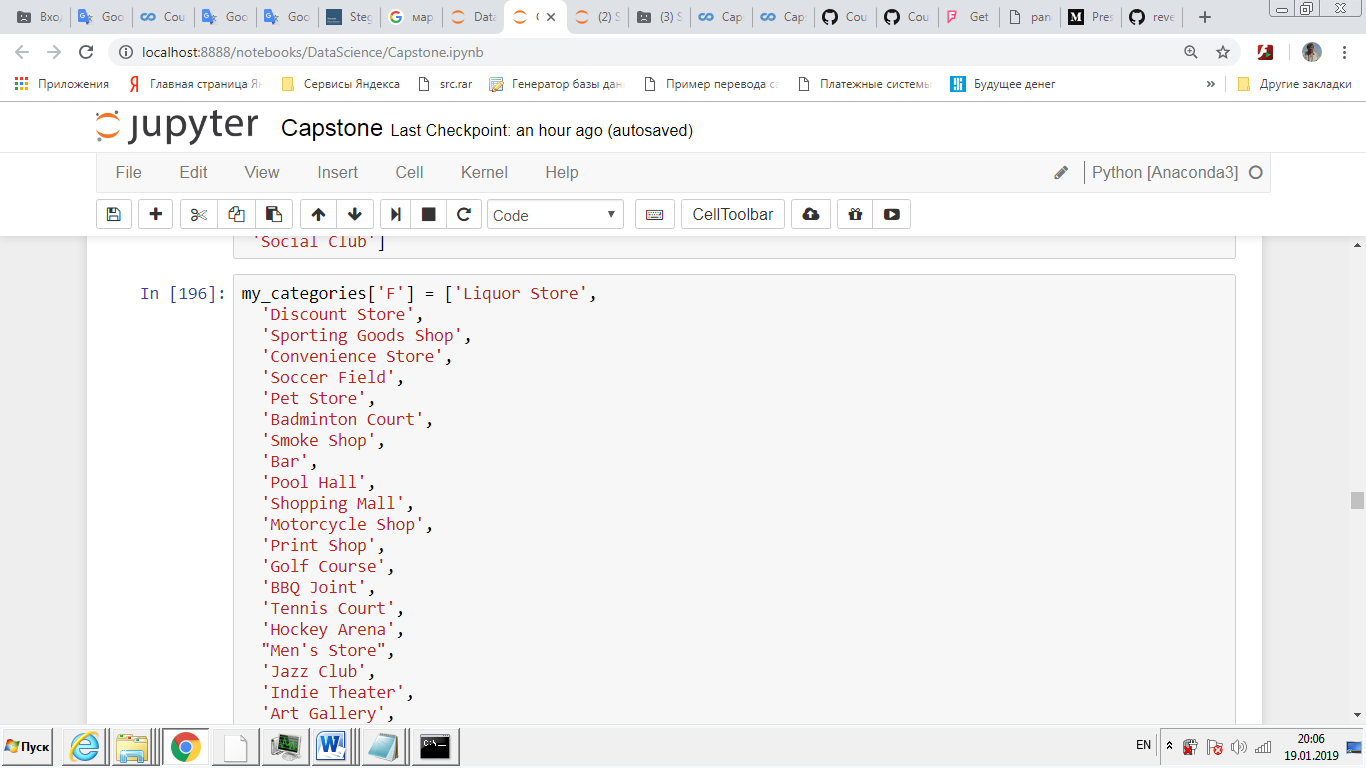
There are very match Venue Category - 331



## 5.5. We will create categories of venues that each family member likes - husband (F), wife (W), daughter (D)

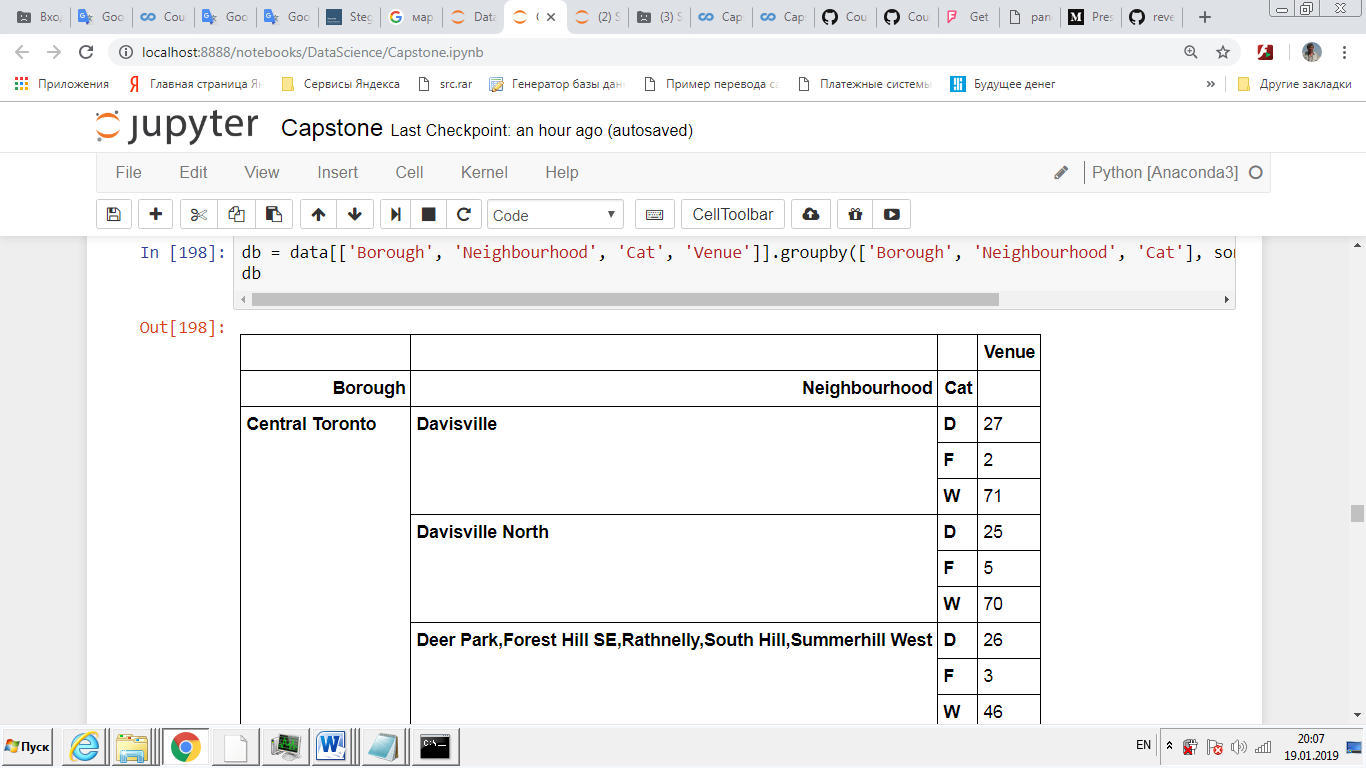




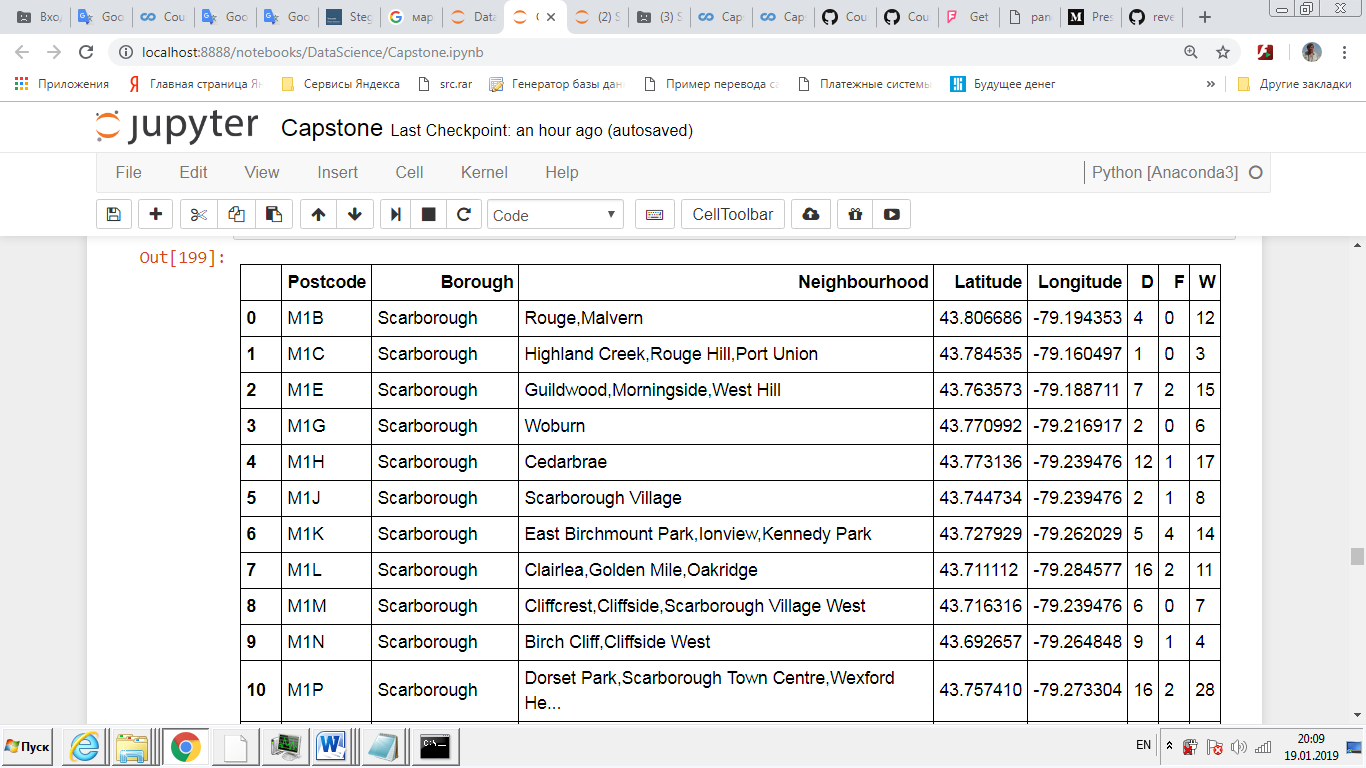


## 5.6. For each venue we add its category (column 'Cat') to 'data' - F, B or D

## 5.7. Find the number of venues that each family member likes in each of the city districts



## 5.8. Add the values found to "df" common data frame

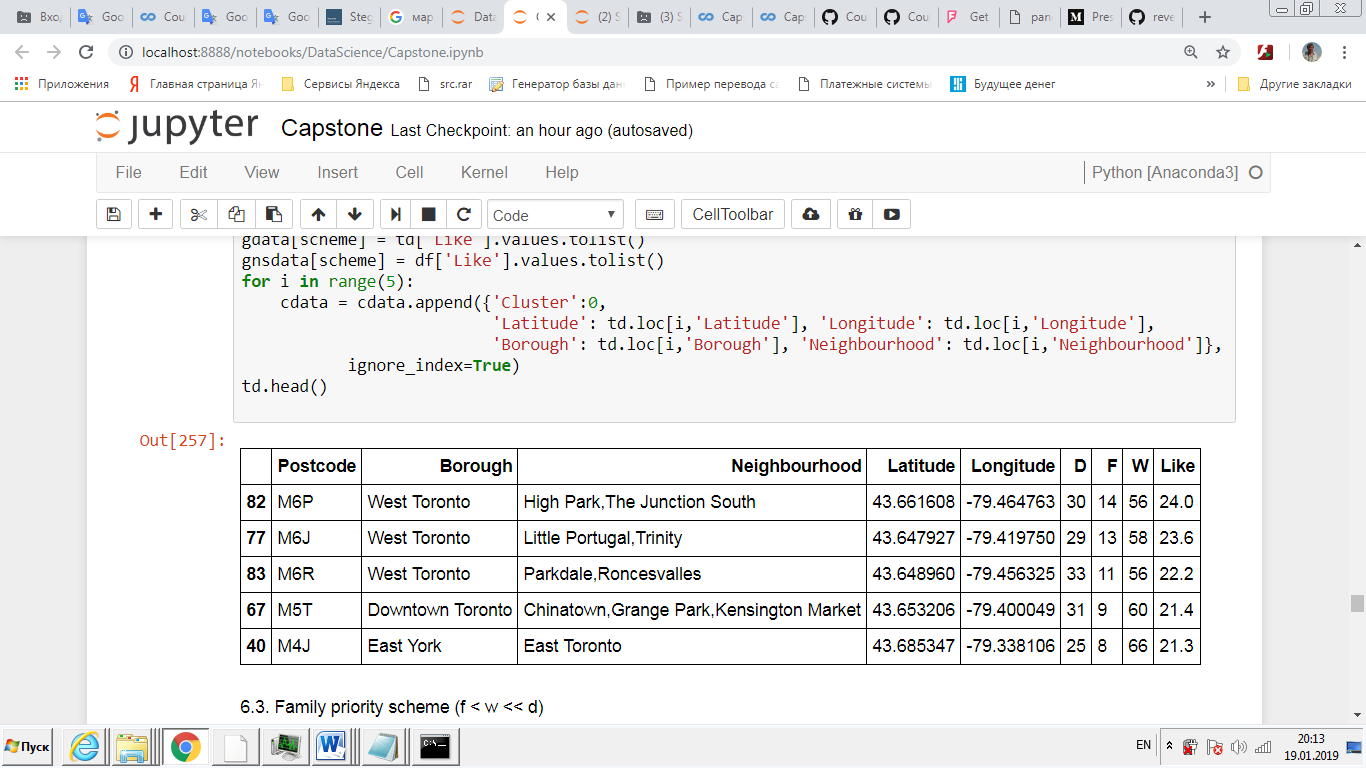


# 6. Results section

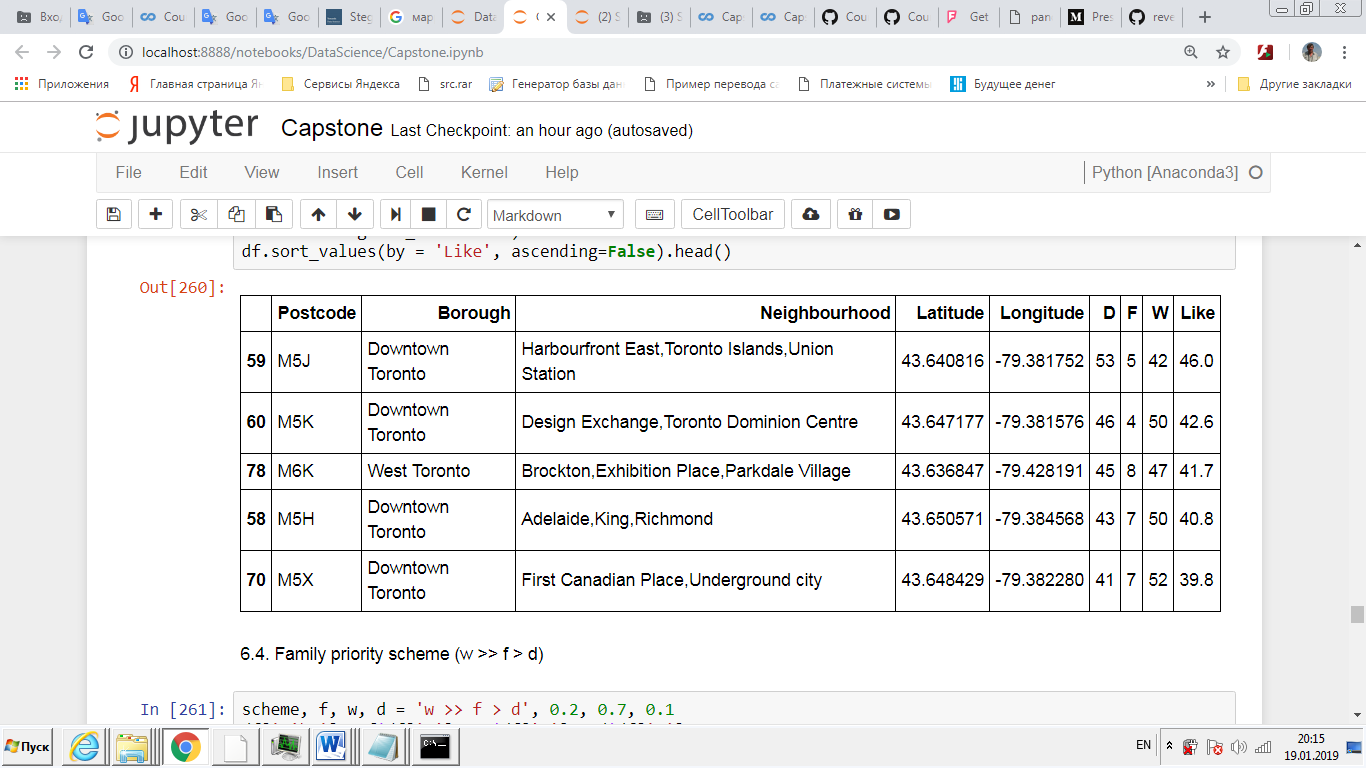
We will find the most pleasant for living areas in Toronto with different family priority schemes.

Who is the most important family member - who is less important - who is the least) and various factors (h, w, d)

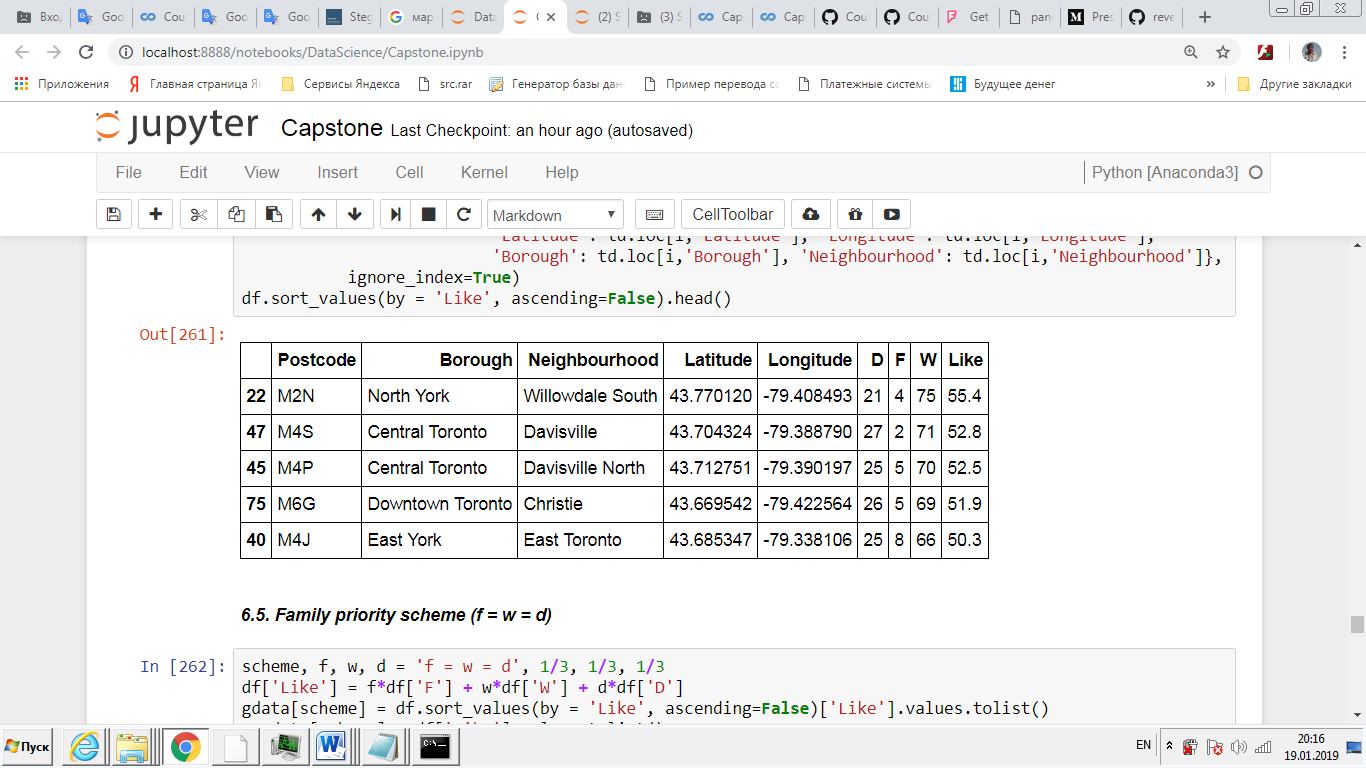
## 6.1. Family priority scheme (f >> w > d) (0.7, 0.2, 0.1)



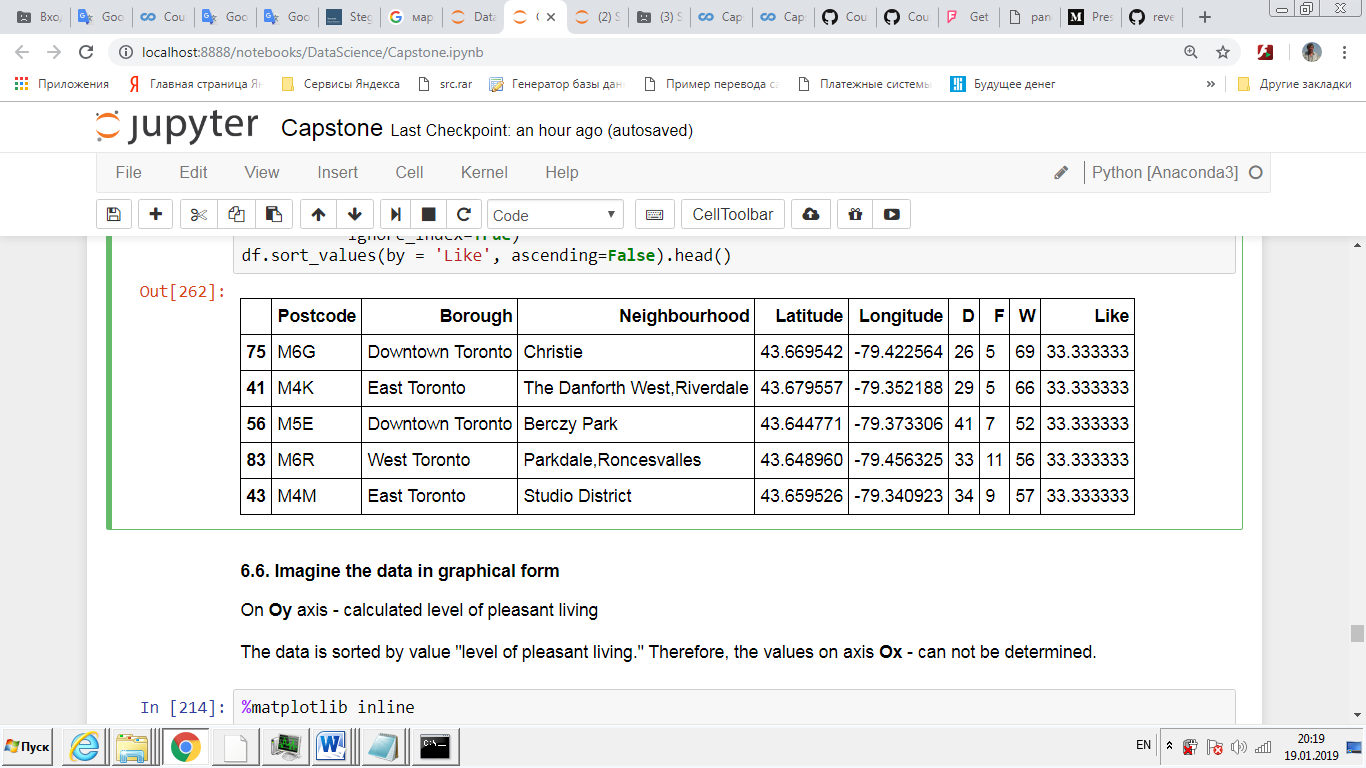
## 6.2. Family priority scheme (f < w << d) (0.1, 0.2, 0.7)



## 6.3. Family priority scheme (w >> f > d) (0.2, 0.7, 0.1)



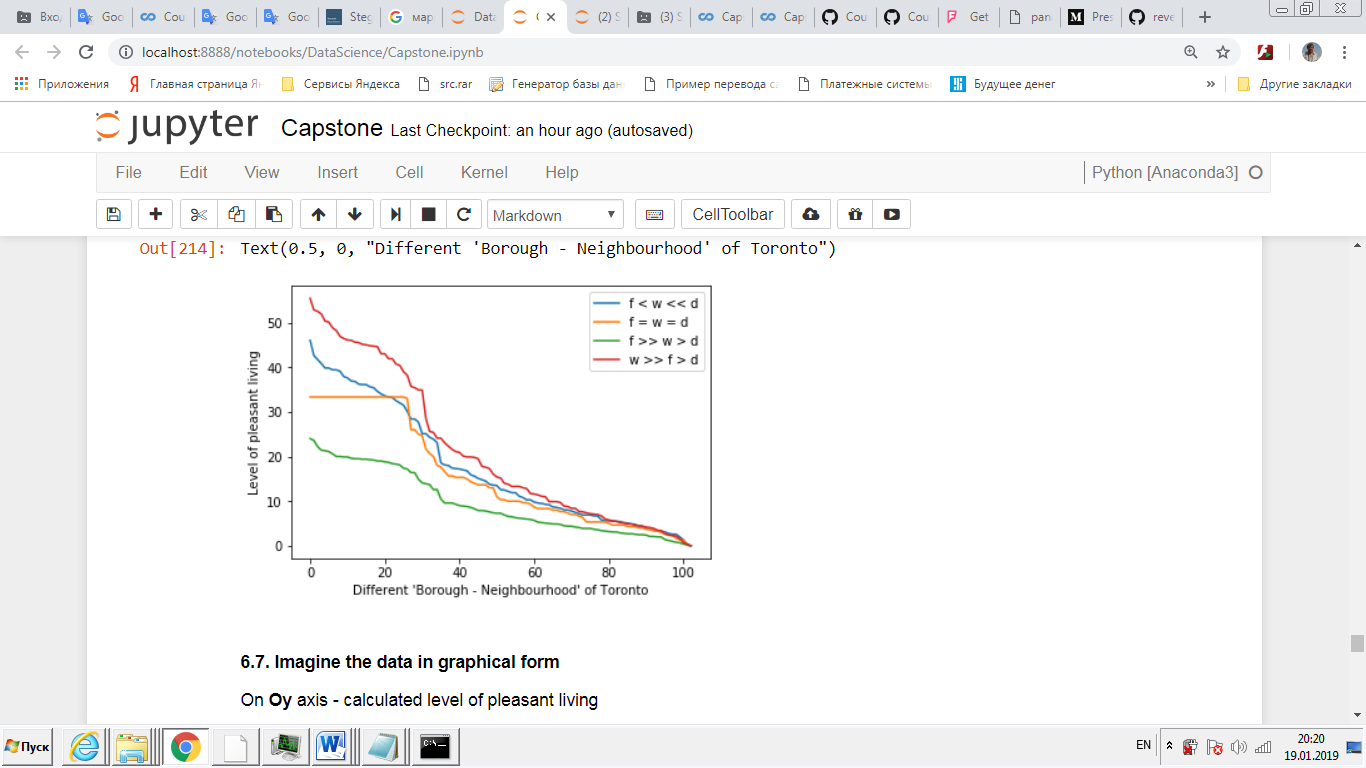
## 6.4. Family priority scheme (f = w = d) (1/3, 1/3, 1/3)



## 6.5. Imagine the data in graphical form

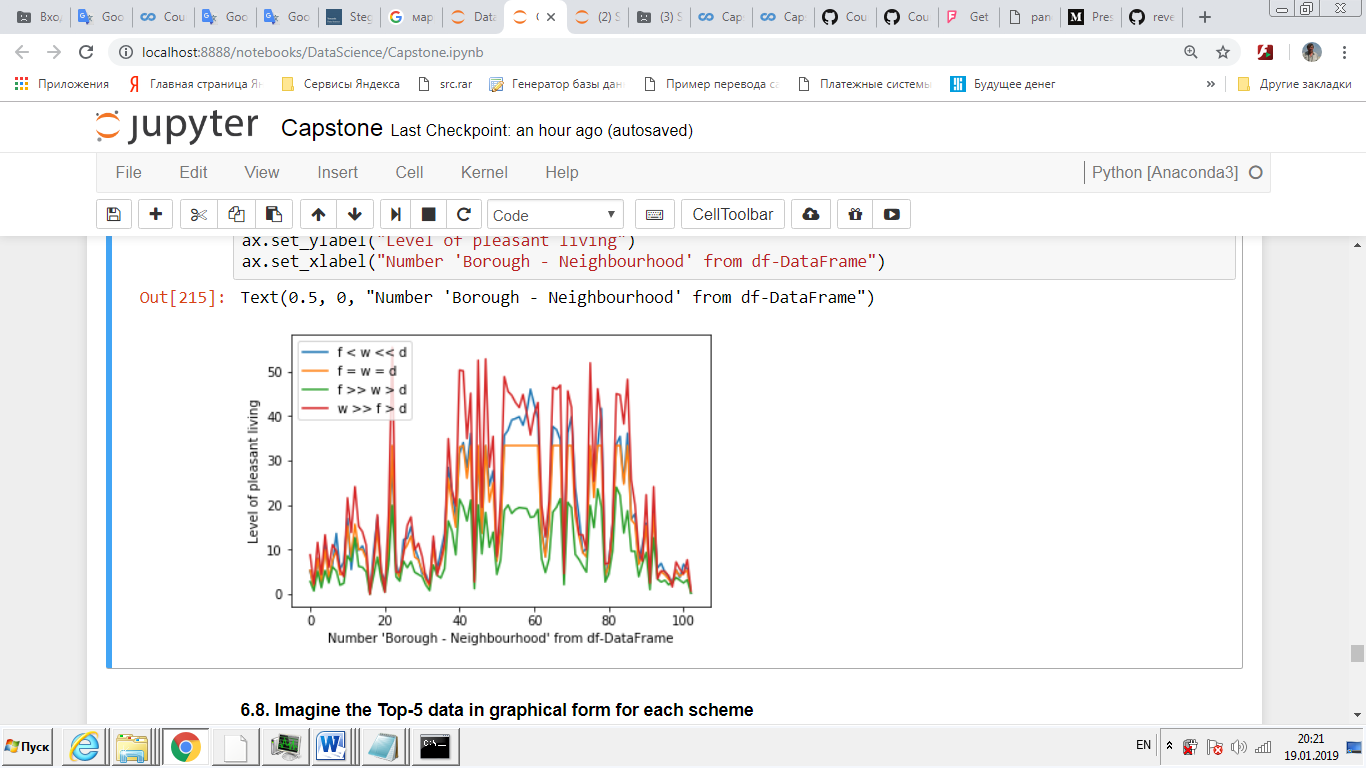
On **Oy** axis - calculated level of pleasant living

The data is sorted by value "level of pleasant living." Therefore, the values on axis **Ox** - can not be determined.



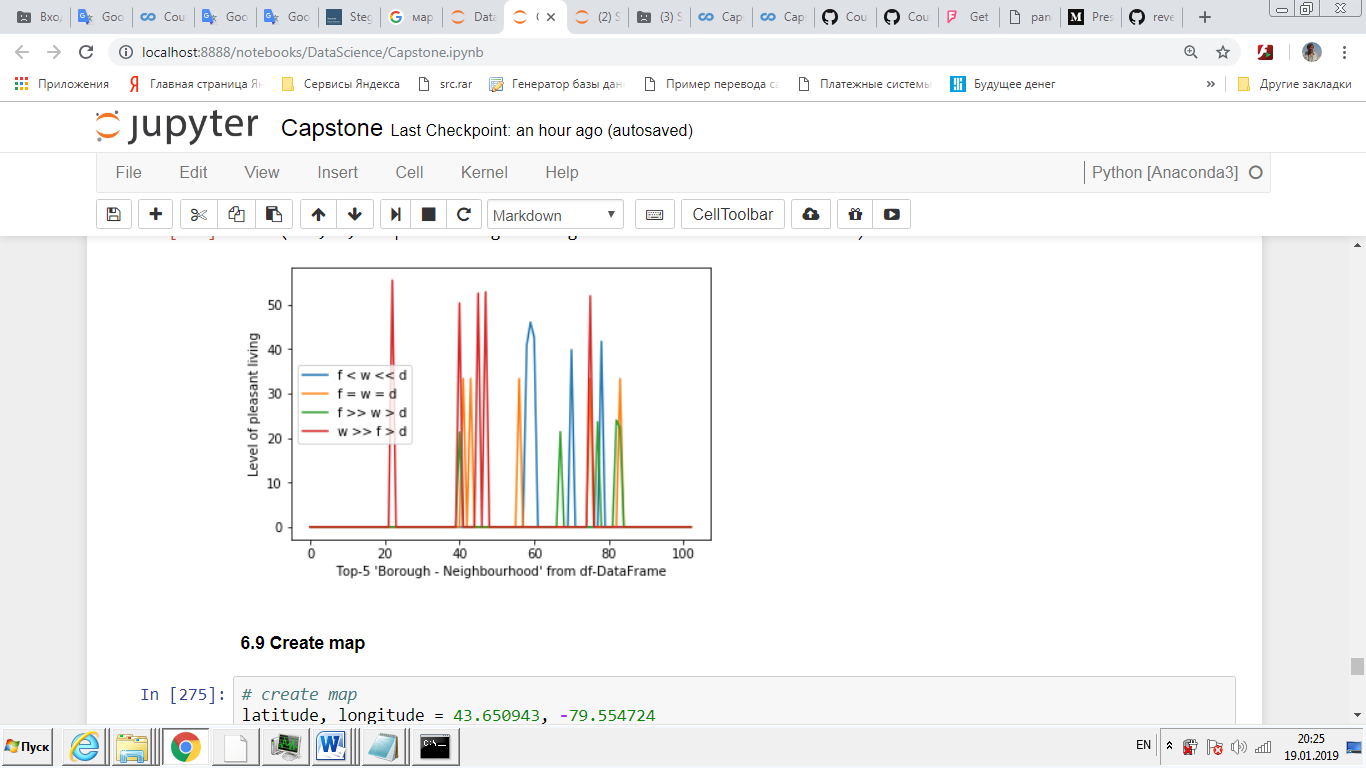
## 6.6. Imagine the data in graphical form

On **Oy** axis - calculated level of pleasant living

The data is not sorted by value "level of pleasant living." Therefore, on axis **Ox** - number of **'Borough - Neighbourhood'**.

## 6.7. Imagine the Top-5 data in graphical form for each scheme

On **Oy** axis - calculated level of pleasant living

The data is sorted by value "level of pleasant living." Therefore, the values on axis **Ox** - can not be determined.

# 7. Discussion section

7.1. As can be seen from graph 6.6, the largest number of places to stay, from which all family members can enjoy, is achieved in the case when the **main person in the family is a wife**.

The **worst case** is that - the main person in the family is **the husband**. In this case, finding a place becomes very difficult.

7.2. As can be seen from graph 6.7, the same area has a completely different level of attractiveness for life from the position of different family members.

Moreover, the requirements of the husband are met the worst in almost all areas.

7.3. As can be seen from graph 6.8, the areas best suited for living from the point of view of various family members do not practically intersect.

Of the 16 districts selected in four schemes, only 4 identical ones were found.

# 8. Conclusion section

Data analysis is a great tool.

It allows you to make interesting discoveries and find unexpected results.