



City life

Taxi Routes and Neighborhoods

*Summary: This project is an introduction to geospatial analytics: GeoPandas, clustering, making maps*

*Version: 1*

# Contents

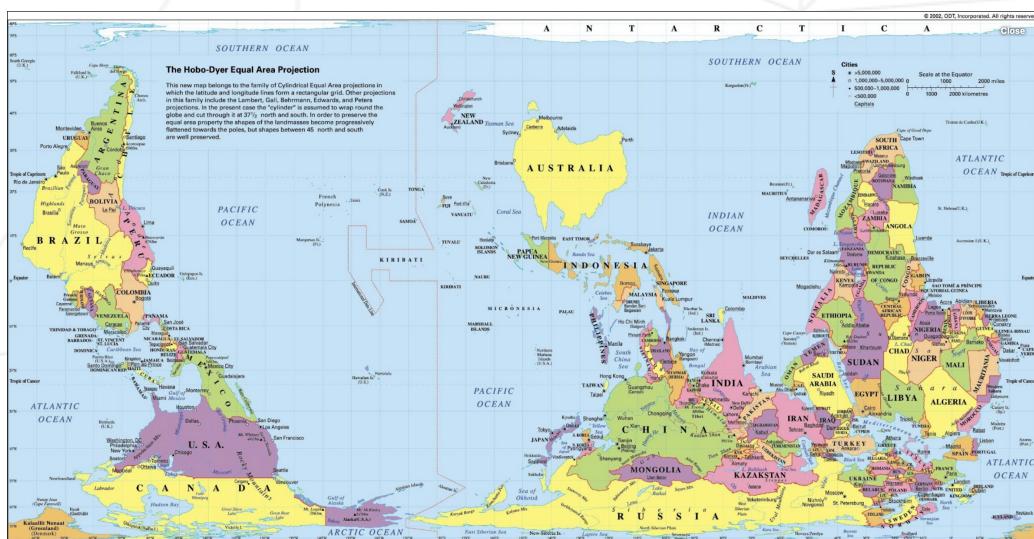
I	Preamble	2
II	Introduction	4
III	Rules of the project	5
IV	Instructions	6
V	Mandatory part	7
VI	Bonus part	10

# Chapter I

## Preamble

Maps are important for humans. They help us better understand what is going on and make decisions. We can see a landscape, we can see different items or objects on it and how they relate to each other in terms of distance or any other term.

Maps have been widely used in battles. It is hard to find anything more useful than them for strategy and tactics creation. When we talk about some space or area where we have to navigate, to draw a map on a napkin is almost a reflective action. It is hard for us to visualize to another person what we see right now by our imagination without a piece of paper and a pencil.



The important thing – the same objects can be mapped in many different ways.  
[\[Source\]](#)

Several decades ago the way how maps exist in our world and the way how we use them changed. We do not use maps in paper form so widely as it had been before. Sometimes to travel from point A to point B in the city required having several maps just in case. They might have had the different quality for different areas and different years of creation as well.

Nowadays maps are in our smartphones. We can easily interact with them. We can search through them. We can calculate the distance between the two spots.

We can get the info about the time required to reach the destination point via different transports. Millions of people who lived in the past would envy.

But maps are not obligatory equal to geography. People can map literally anything: sky, DNA, network topology, business environment, brain, galaxies, buildings, engineering systems.

# Chapter II

## Introduction

Geospatial analytics is a field that appears in the intersection of data analysis and data visualization. It is widely used in telecom, logistic, transport, the oil industry, agriculture companies, and government agencies that manage territories. Any time when you need to get insights from data that have coordinates in space, geospatial analytics is a must.

The most common tasks are different but almost all of them are connected to maps. For example, the list of responsibilities from a job description:

- Analyzing spatial data through the use of mapping software.
- Discovering patterns and trends through spatial mapping of data.
- Designing digital maps with geographic data and other data sources.
- Creating "shapefiles" to merge topographical data with external data by layering external data over a topographical map.
- Producing maps showing the spatial distribution of various kinds of data, including crime statistics and hospital locations.
- Developing mapping applications and tools.
- Converting physical maps into a digital form for computer usage.
- Performing data munging and cleaning to convert data into its desired form.
- Produce reports on geographic data utilizing data visualizations.
- Managing a digital library of geographic maps in various file types

So there are just a few special algorithms for geospatial data, and most of them are visualizing data on maps.

# **Chapter III**

## **Rules of the project**

The goal of this project is to give you a first approach to Geospatial Analytics. You will try different tasks to analyze taxi rides. The efficiency of any taxi business depends on it.

# **Chapter IV**

## **Instructions**

- This project will only be evaluated by humans. You are free to organize and name your files as you desire.
- Here and further we use Python 3 as the only correct version of Python.
- The norm is not applied to this project. Nevertheless, you are asked to be clear and structured in the conception of your source code.
- Store the datasets in the subfolder data.

# Chapter V

## Mandatory part

### a. Task

In this project, you will try to get some valuable insights about customers' and taxi drivers' behavior. Maybe it will help a taxi company optimize its business.

Here are the tasks that you need to do:

1. find out and visualize on a map most popular areas where people ordered a taxi as well as where they headed to,
2. find out and visualize the most popular routes in different time intervals,
3. find in the dataset locations of the city infrastructure and visualize how the customers were arriving at one of them using an animated map,
4. visualize one day of a taxi driver and how much money he or she earned using an animated map,
5. visualize one day of the city (working day and weekend day) using an animated map.

### b. Dataset

You will work with another part of the same dataset of taxi drives. It contains data from 2014-09-05 to 2019-06-23. This time it contains data about taxi drivers, coordinates of pick point and drop off locations, timestamps, and fares.

Also, you will find there the files needed to visualize a map of the city.



You can find the dataset in the project page

## c. Implementation

You can work in [Google Colab](#) or Jupyter Notebooks on your computer.

You can use any library or any framework that you find convenient.

### Most popular areas

1. Conduct clustering analysis of pick point and drop off locations based on their coordinates. Clusters might be different for each of the categories (pickpoints and drop-offs).
2. Draw the borders and centroids of the clusters on a map. You will have two maps.
3. Use a color scale to show which clusters (i.e. areas) have the largest number of pick points and drop-offs.

### Most popular routes

1. Conduct a clustering analysis of taxi rides based on the coordinates of the pick point location and drop off.
2. Draw centroids of the top 5 popular clusters (routes).
3. Draw routes between centroids of a cluster – not a direct line from point A to point B, but a path that takes into account the city streets.

### City infrastructure

1. Find locations of the city infrastructure (airports, stadiums, parks, universities) using the data and make up your own algorithm of how to find them. Find at least 6 of that kind of location.
2. Find the rush hour for each of the locations – timestamp when the location had the largest number of dropoffs and save the information into that file : "rush\_hours\_empty.csv" (in attachment).
3. Visualize one day of any of the locations including that rush hour showing how people from different places were coming to the location and then were leaving it on an animated map.
4. When the ride finishes, it should disappear on the map. In other words, it should look like firing neurons in the brain.
5. The map should display time.

## One day of a taxi driver

1. For the taxi driver with ID ([2ea4ad2950f3bbdfdcfa7adb48e0dcee49d8a714b7024342f0302  
eeb9e891dfd55a6f35bb7bc7af06398fb4f55583e1659cb11b432848296bfd2b7d3084e7de1](#))  
visualize his or her rides during the day (2019-05-31). Display the current amount  
of money earned.
2. When the ride finishes, it should NOT disappear on the animated map.
3. Every time the ride finishes the counter of money should be updated.
4. The map should display time.

## One day of the city

1. Visualize all the rides in the city during the day (2019-05-16) on an animated map.
2. When the ride finishes, it should disappear on the map. In other words, it should  
look like firing neurons in the brain.
3. The map should display time.

## d. Submission

Your repository should contain one or several notebooks with your solutions and visualizations.

Also, it should contain the file with rush hours for the city infrastructure objects. It will be compared to our file. We will calculate the intersection of your set with our set of locations and rush hours. The intersection should have at least 3 elements. The columns that will be taken into account: longitude, latitude, num\_of\_rides, Trip End Timestamp.

# **Chapter VI**

## **Bonus part**

- Design your visualizations as a website with interactive elements (for example, where you can choose a date or driver id, choose a location, etc).
- Try to achieve a better result with the intersection – at least 5 elements.