# Project Introduction

Cordis (a Cardinal Health company) develops and manufactures medical devices for medicine and interventional procedures to treat patients littered with coronary and peripheral vascular diseases.   
In order to expand it’s business everywhere the world, Cordis company began to invest in many various comes that they suppose are going to be valuable for them within the close to future. each investment they created, each funding they funded is recorded during this dataset.  
This dataset contains all comes funded by the eu Union beneath the fifth framework programme for analysis and technological development (FP5) from 1998 to 2002.  
The file 'FP5 Projects' contains the general public grant info for every project, together with the subsequent information: Record management range (RCN), project ID (grant agreement number), project form, project standing, funding programme, topic, project title, project begin date, project finish date, project objective, project total price, international organisation goop contribution (commitment), call ID, funding theme (type of action), arranger, arranger country, participants (ordered in an exceedingly semi-colon separated list), participant countries (ordered in an exceedingly semi-colon separated list).  
The collaborating organisations area unit listed within the file 'FP5 Organisations' that includes: project Record management range (RCN), project ID, project form, organisation role, organisation ID, organisation name, organisation short name, organisation kind, participation terminated (true/false), international organisation contribution, organisation country.

# Analysis OBjectives

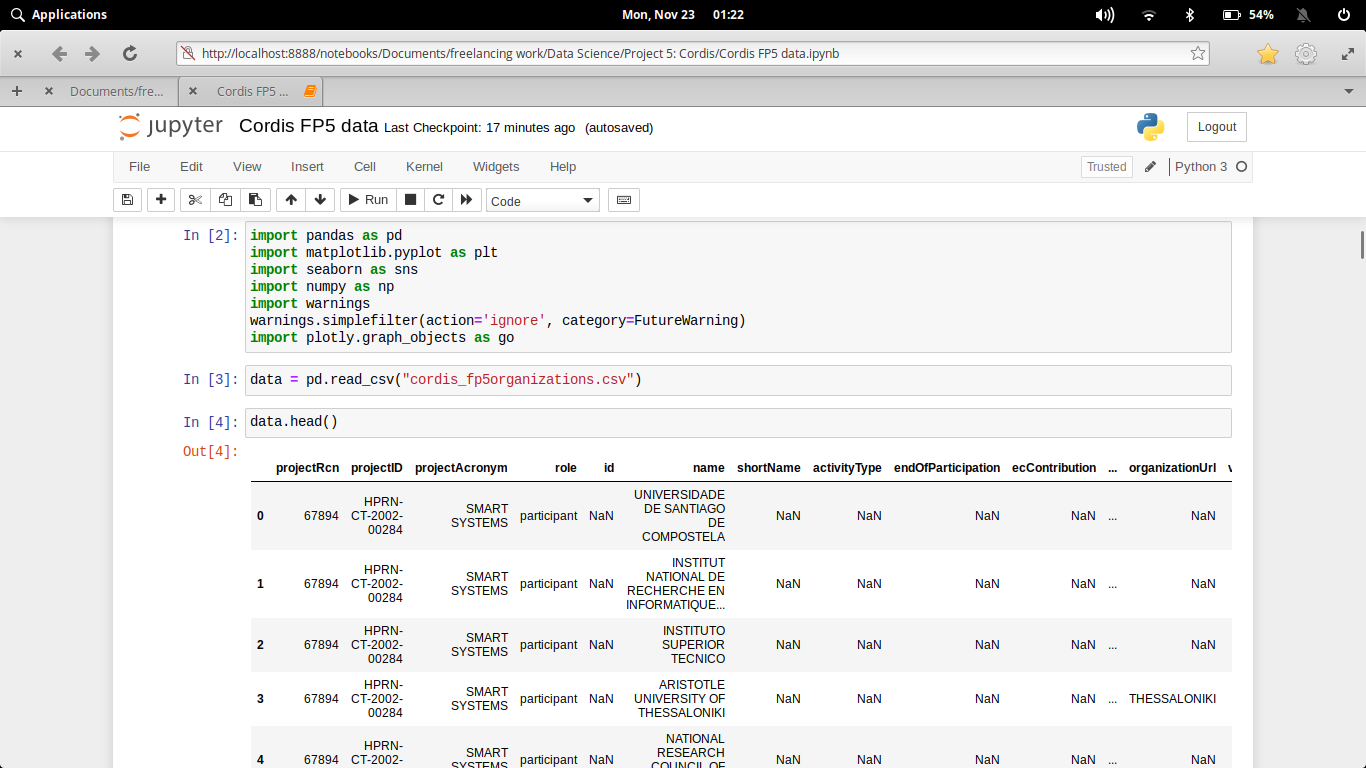
Here are the list of question whose analysis which be given below..

1. How many different types of roles are there and what are the top roles?
2. What is the coordinator to participant ratio?
3. Which are the top 20 countries with maximum investment?
4. Which top 20 countries contributed from maximum number of states?
5. Which locality(street) is the major contributor in the cordis growth?
6. What is the max number of projects approved with a single identity?
7. What are the some top projectAcronym approved by FP5

# Data Acquisition and Cleaning

**Code to read the data from Excel / CSV / HTML.**

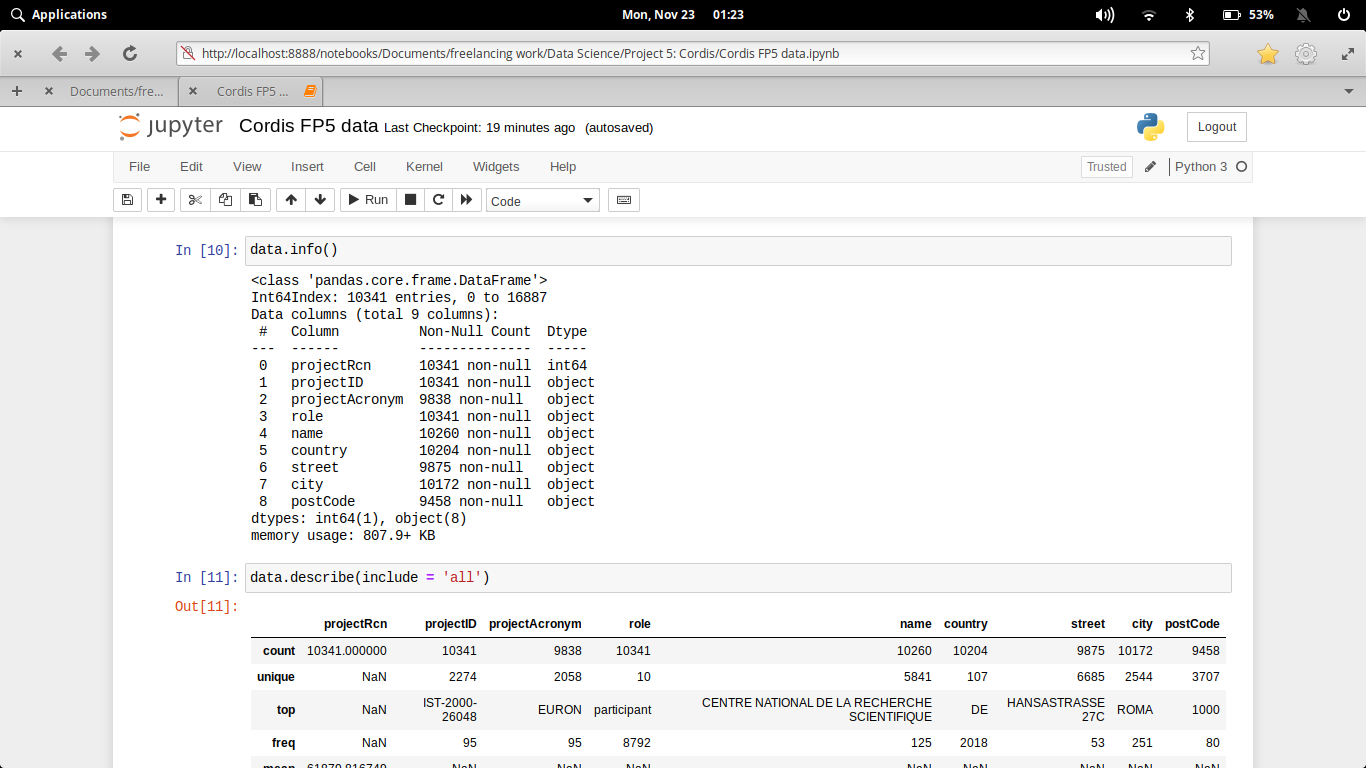
To read the dataset in csv format, we will load it into Pandas data frame but first let’s import the pandas library and set an alias by typing **“import pandas as pd”**. After importing the library with the alias **“pd”**, let us load the .csv file using the following line of code:

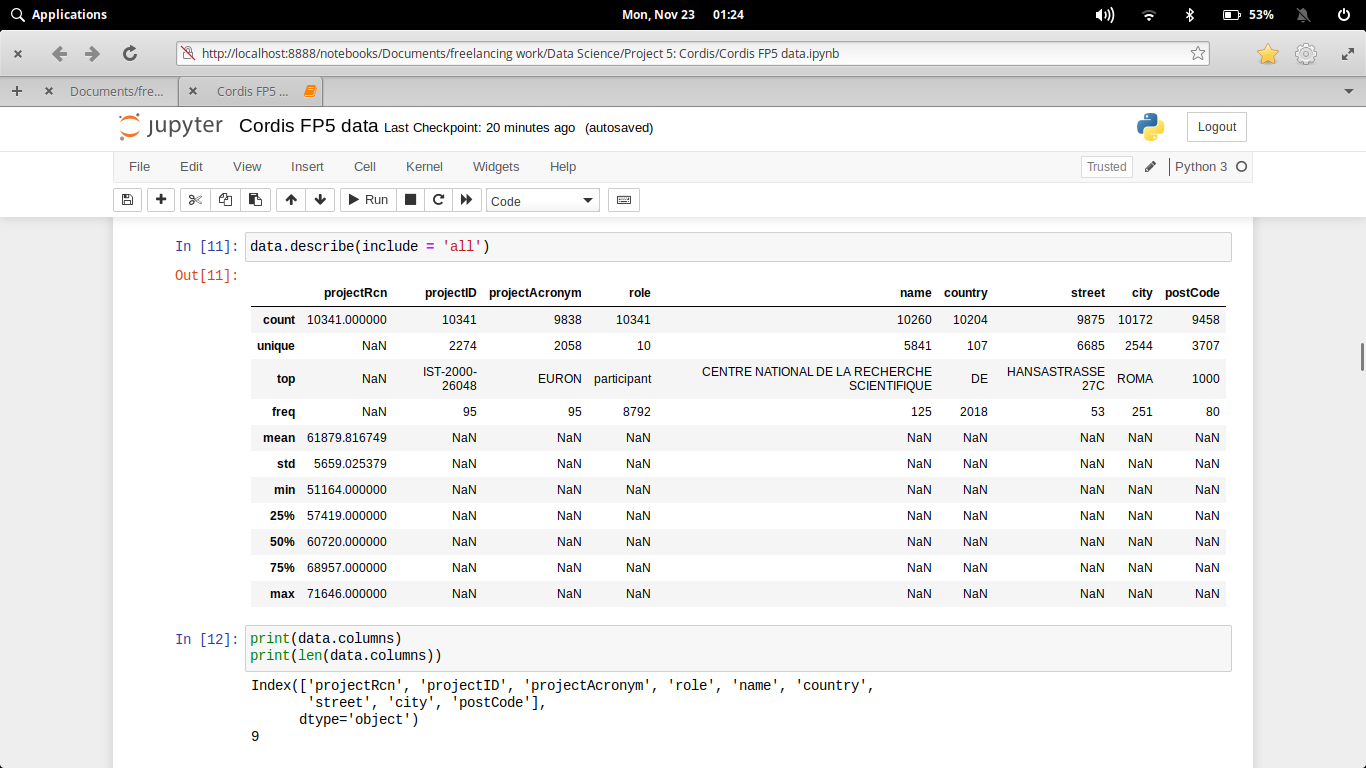


Here we have import our csv files and read through pandas library.

Here the csv file can be read through (Pandas library) and store in **data Dataframe**. The Dataframe can be shown through **.head()**. The number of rows we want to show, that number we have to pass in head parentheses as an argument.

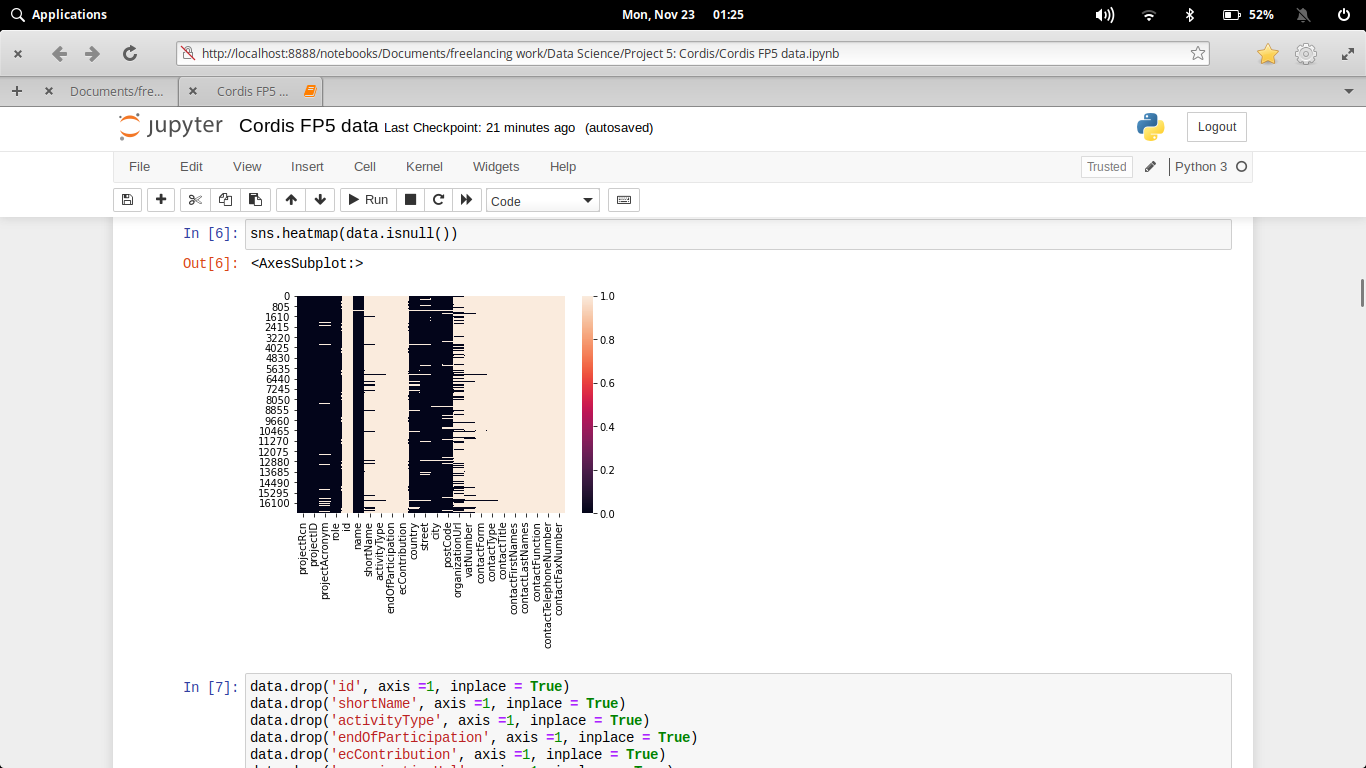
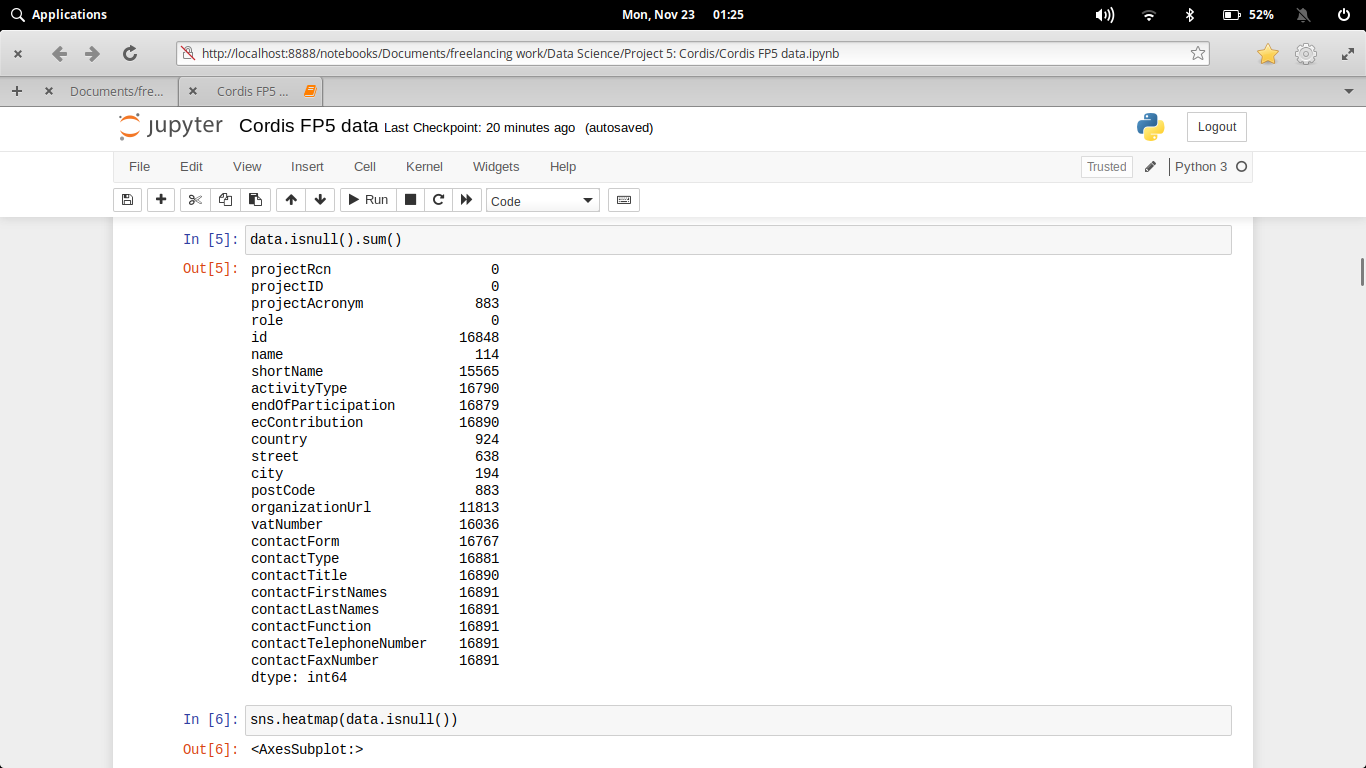
Now if we want to describe our dataframe for our better understanding to know the stats. and other parameter that our dataset should follow.



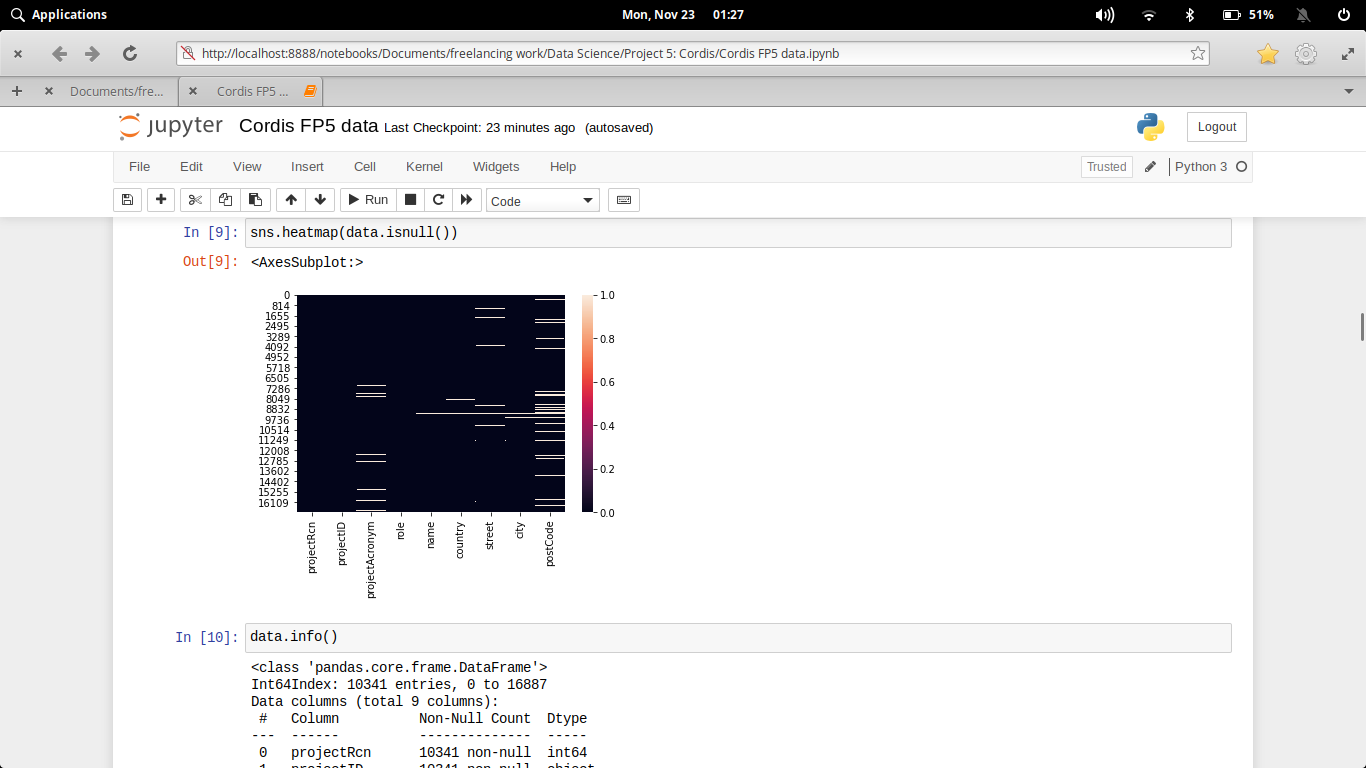
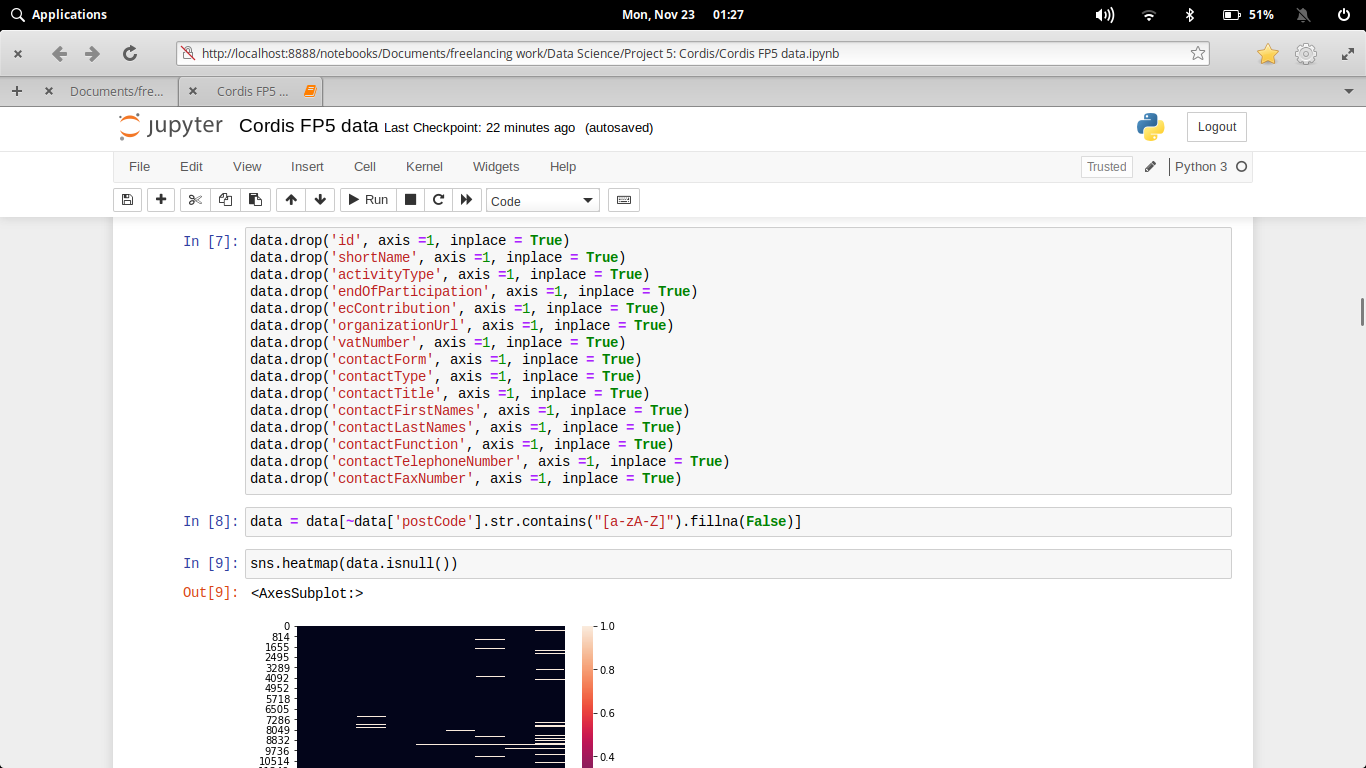


**Clean the unnecessary data, by removing, replace the missing data and renaming the columns.**

Dataset generally contains some null value, which is generally caused by misplacing some values. So its necessary to clean this mess from our dataset for better visualization



In the above diagram we see that some of the columns of this dataset contains null value.



In the process of data cleaning we first drop the particular columns, which contains atmost null value. Then subsequently we drop the rows which contains alphabets as their **postCode** so that the values will not much more effect. Finally after all process our heatmap diagram shows less blank spaces which indicated that our dataset is almost clean.

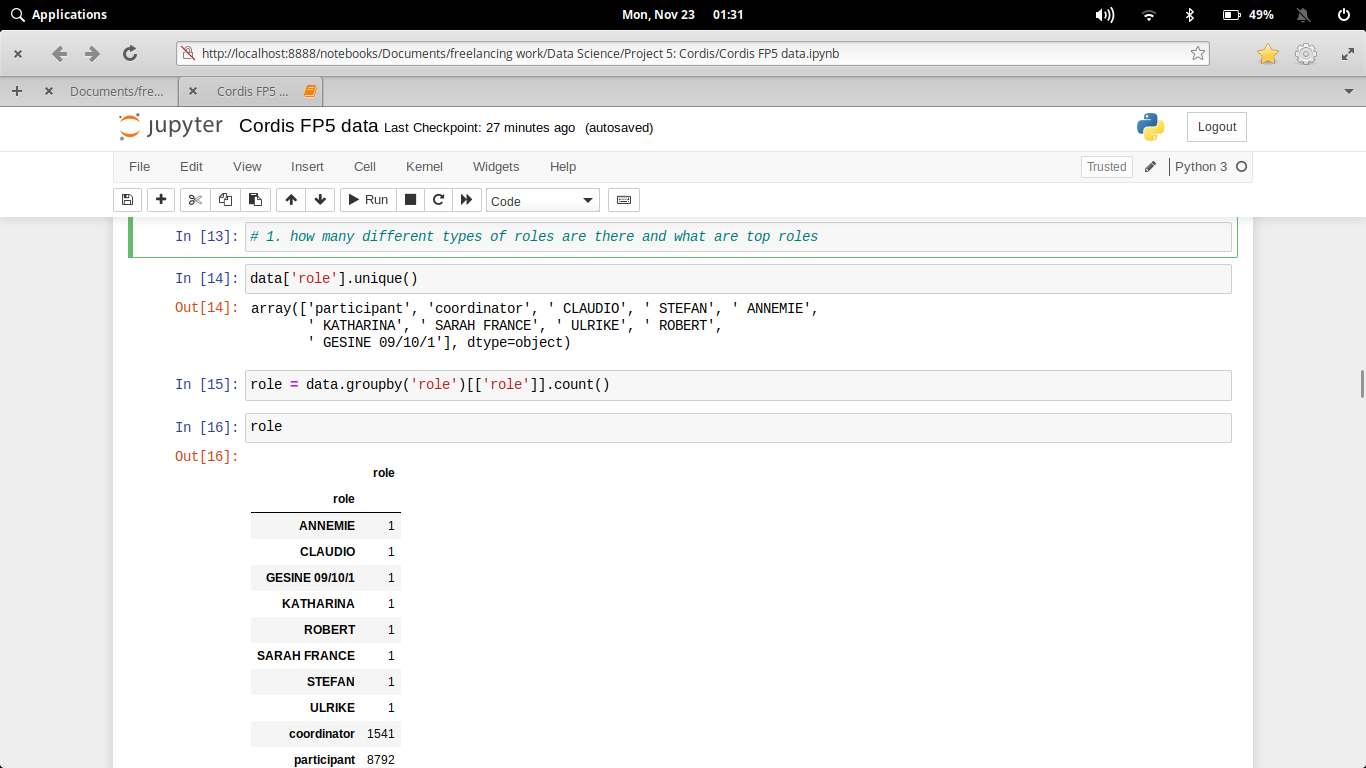
**why data clean needed (for your data)**

Data cleansing or scrubbing or appending is the procedure of correcting or removing inaccurate and corrupt data. This process is crucial and emphasized because wrong data can drive a business to wrong decisions, conclusions, and poor analysis, especially if the huge quantities of big data are into the picture.

# Data and Exploratory Analysis

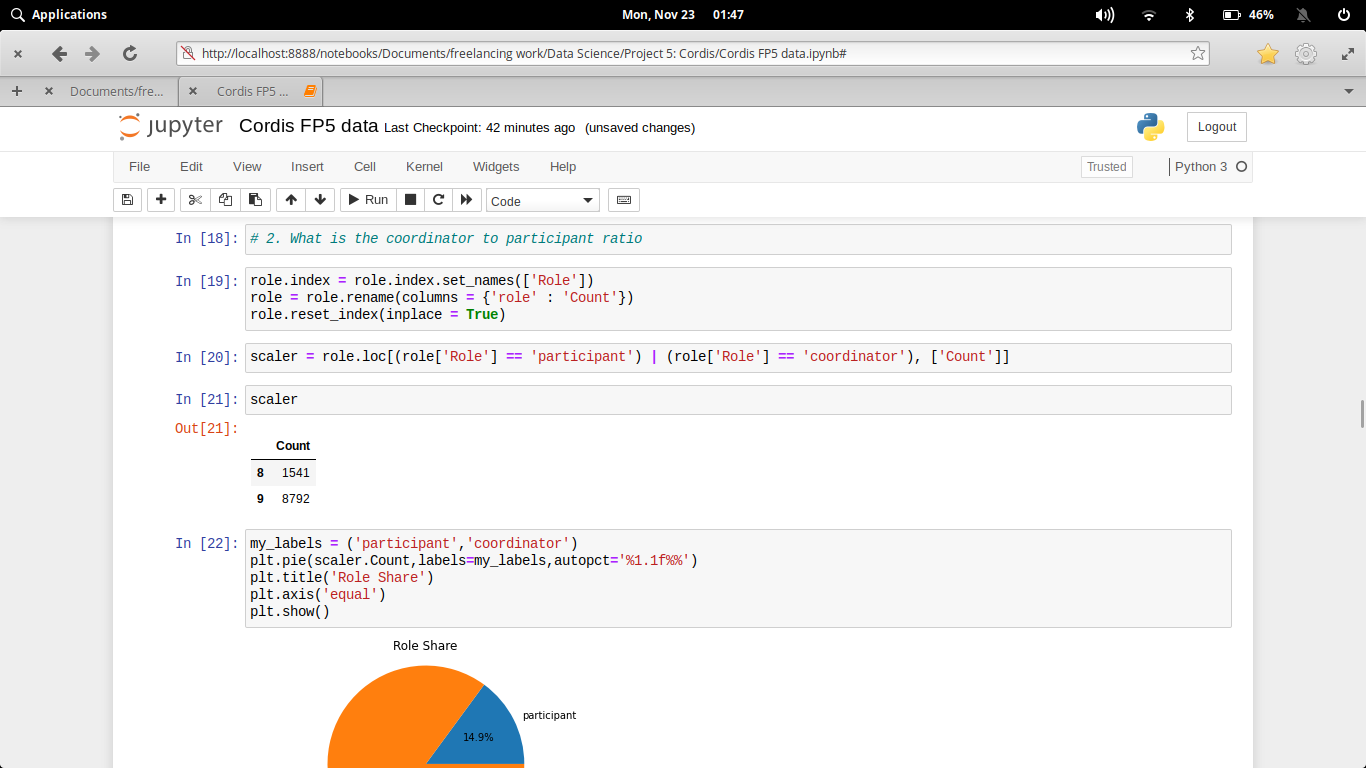
**Code and its output with Explanation**

1. **How many different types of roles are there and what are the top roles?**

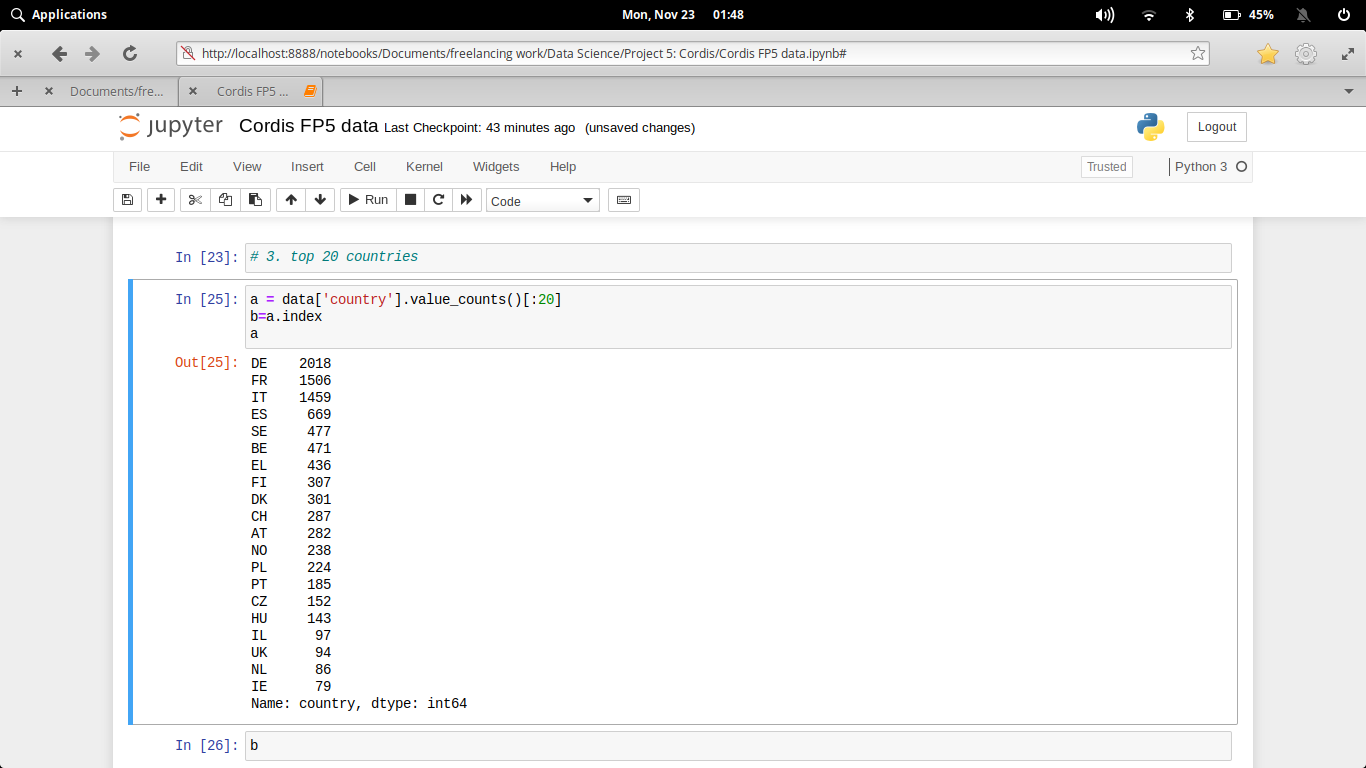
****

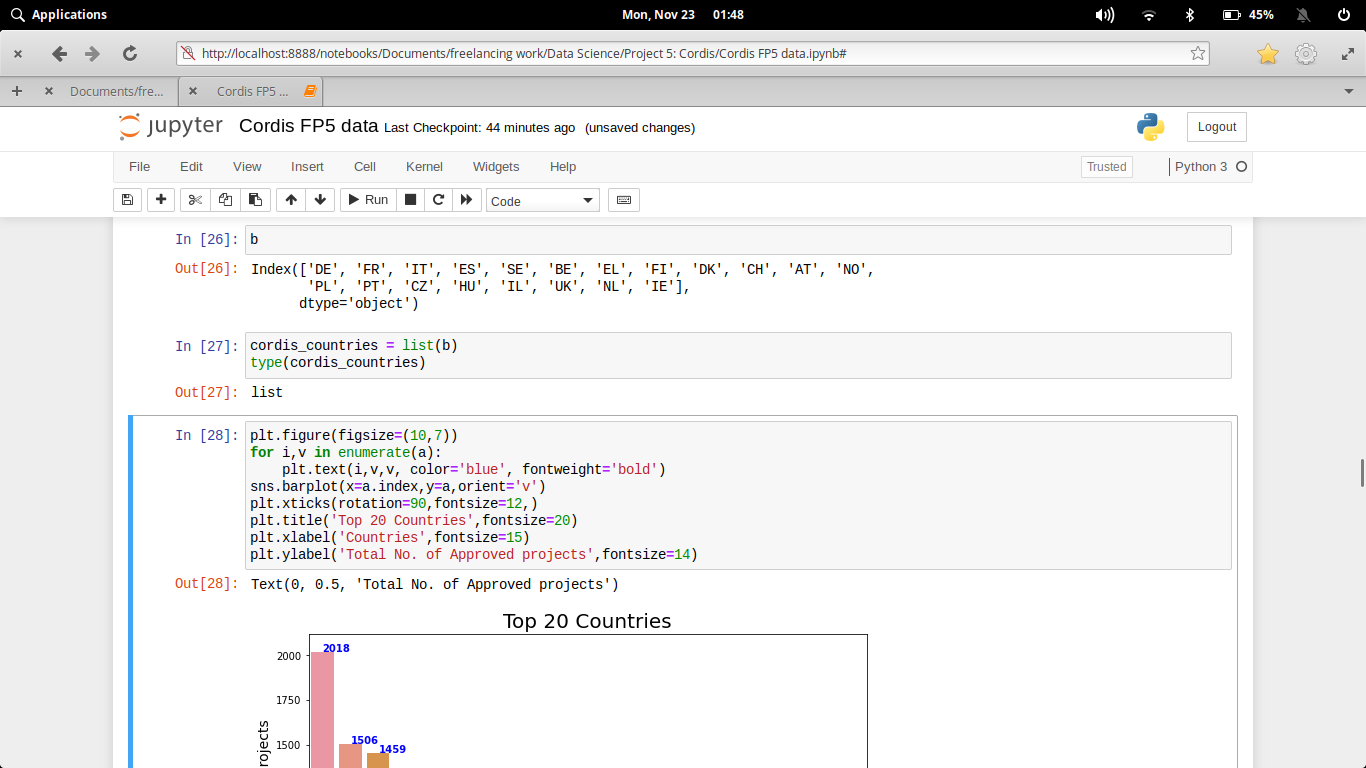
This show that apart from coordinate and participants, there are some more roles which are basically just a name, so, we have to drop those rows too as a part of cleaning process.

1. **What is the coordinator to participant ratio?**

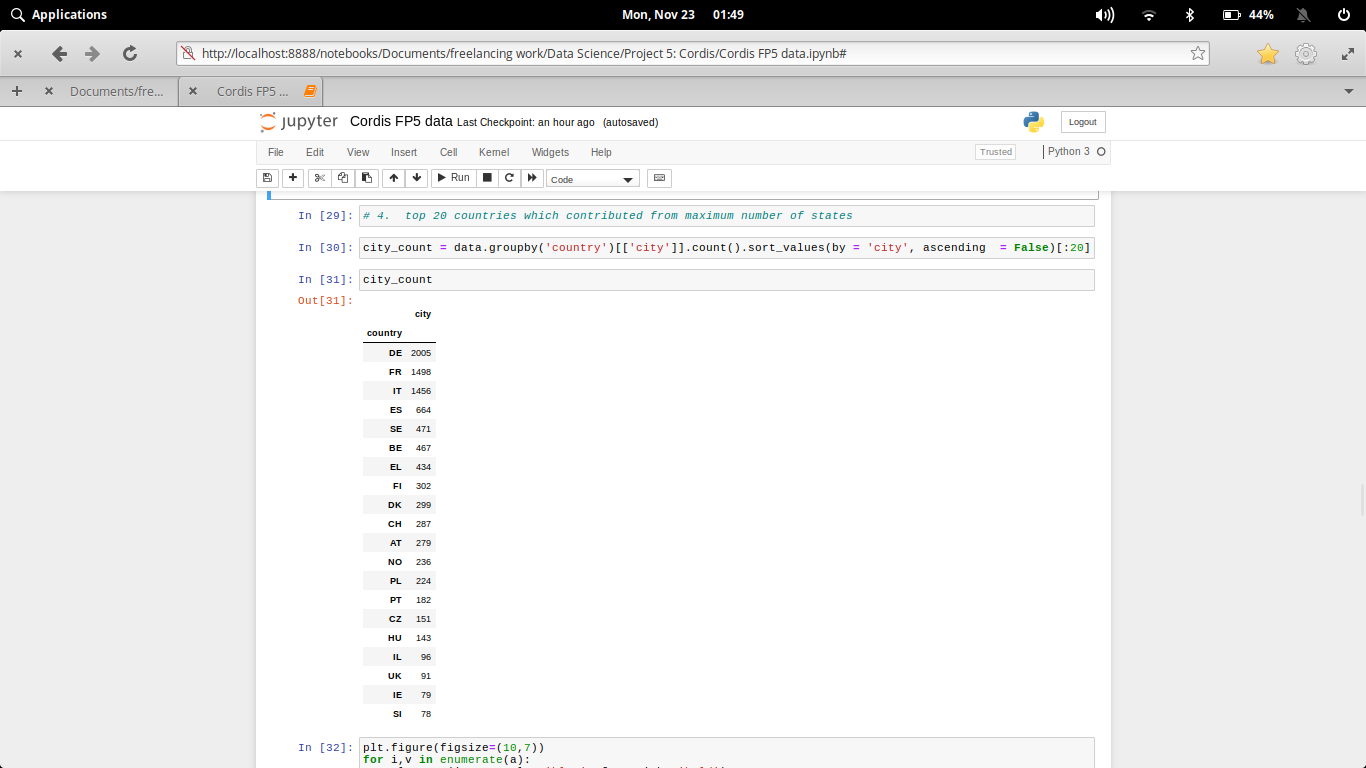
****

1. **Which are the top 20 countries with maximum investment?**

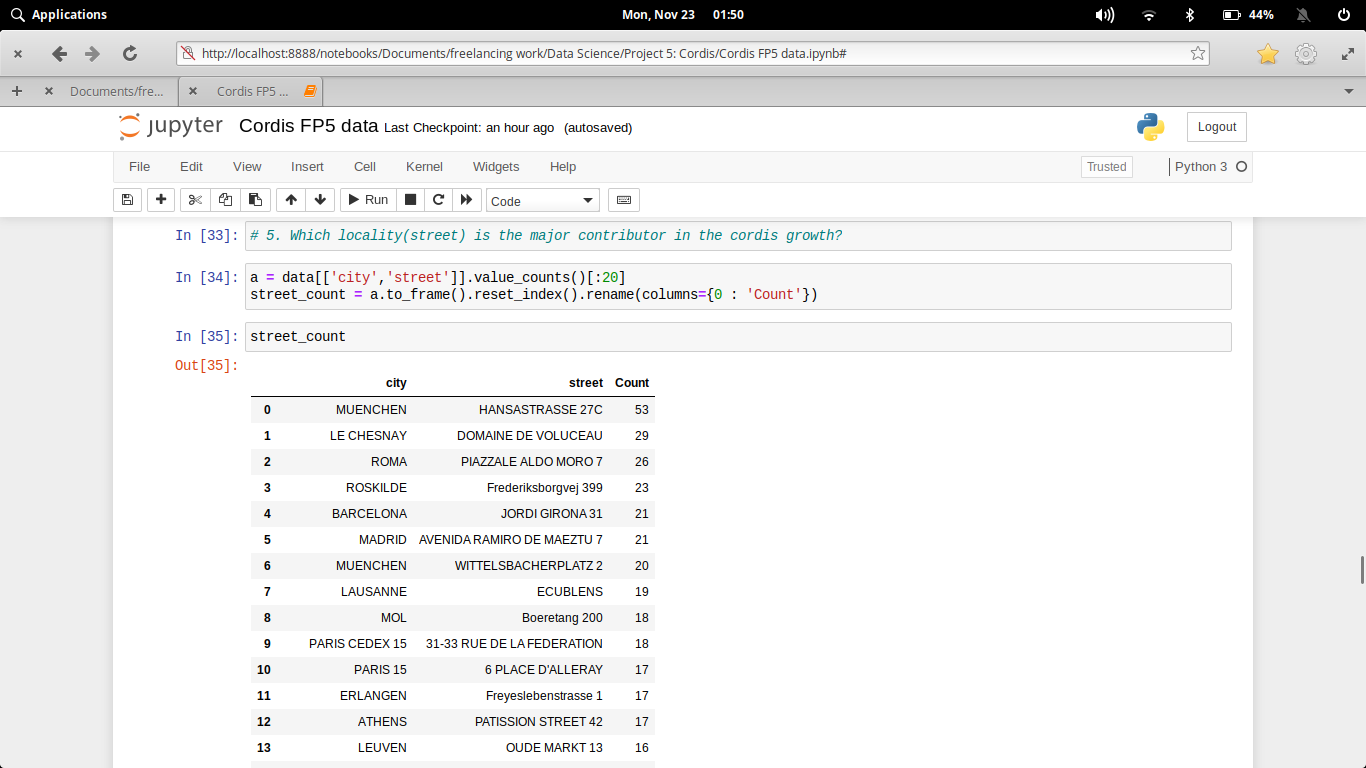
****

****

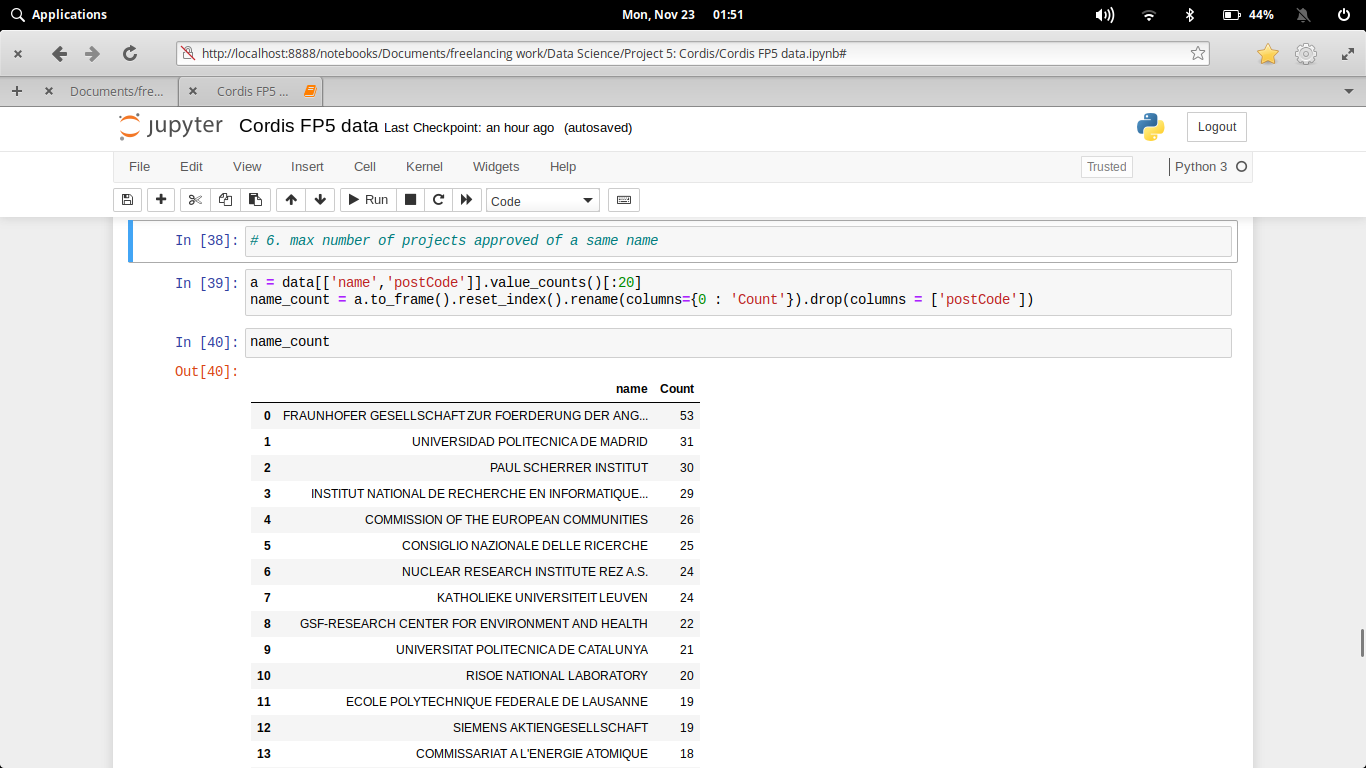
1. **Which top 20 countries contributed from maximum number of states?**

****

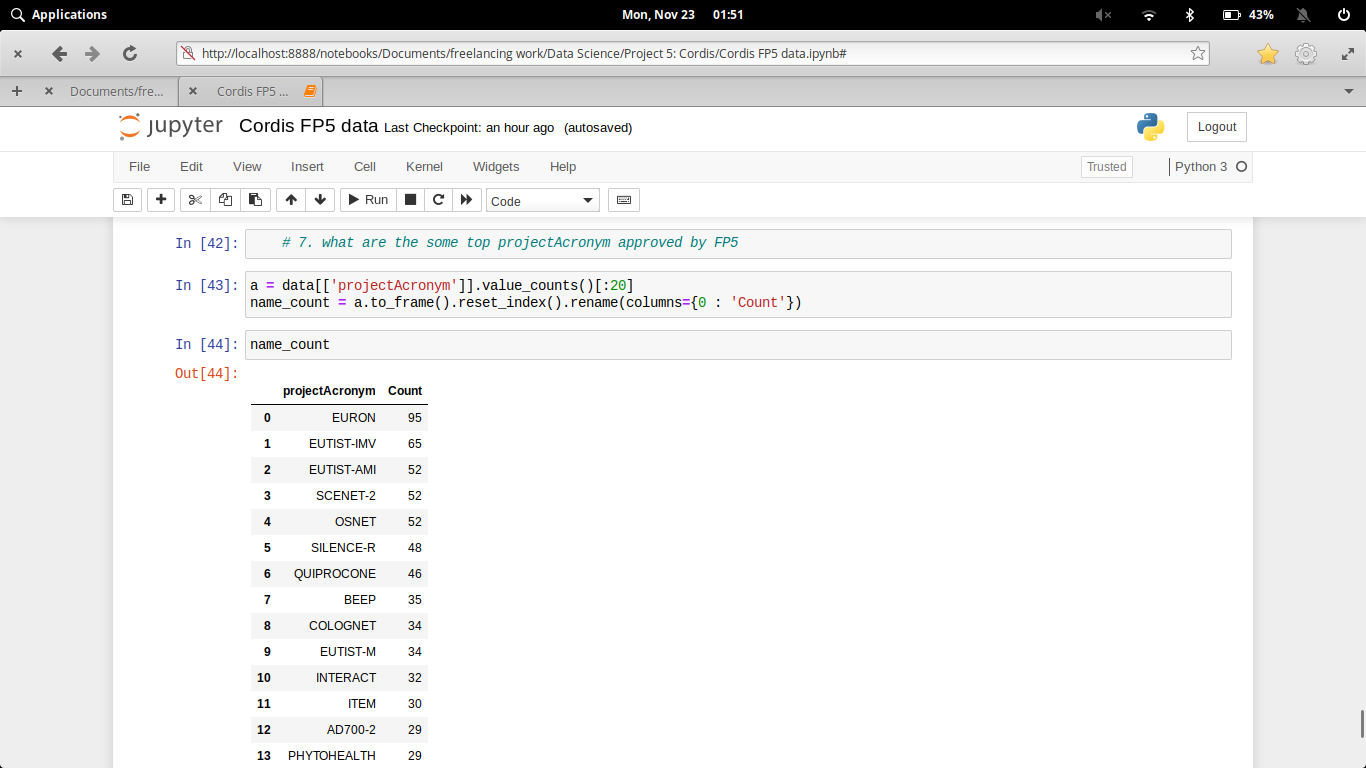
1. **Which locality(street) is the major contributor in the cordis growth?**

****

1. **What is the max number of projects approved with a single identity?**

****

1. **What are the some top projectAcronym approved by FP5**

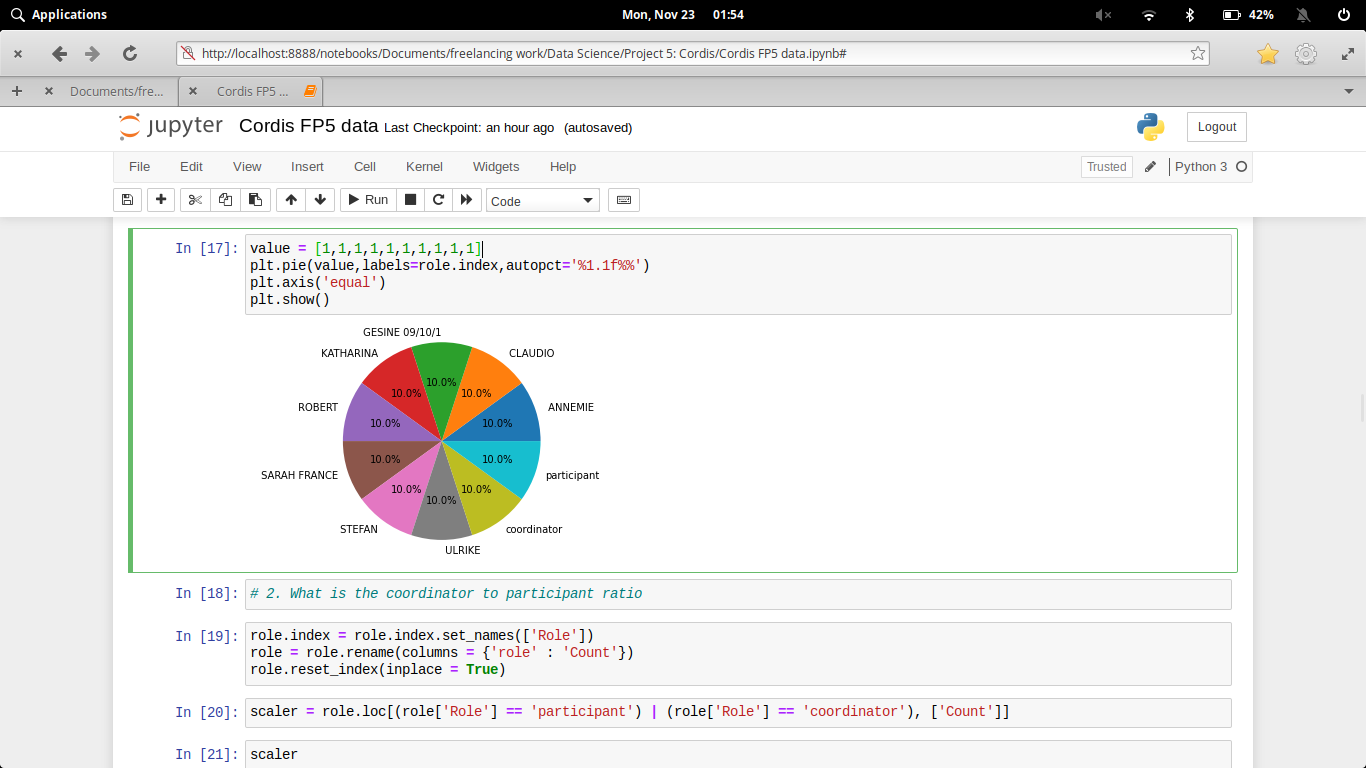


# Data Analysis – Visualization

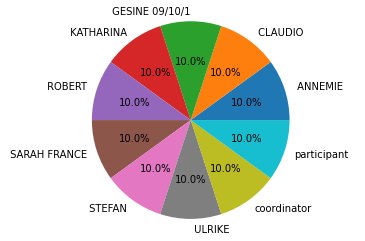
**Code and its output with visualization**

1. **How many different types of roles are there and what are the top roles?**

**Code:-**

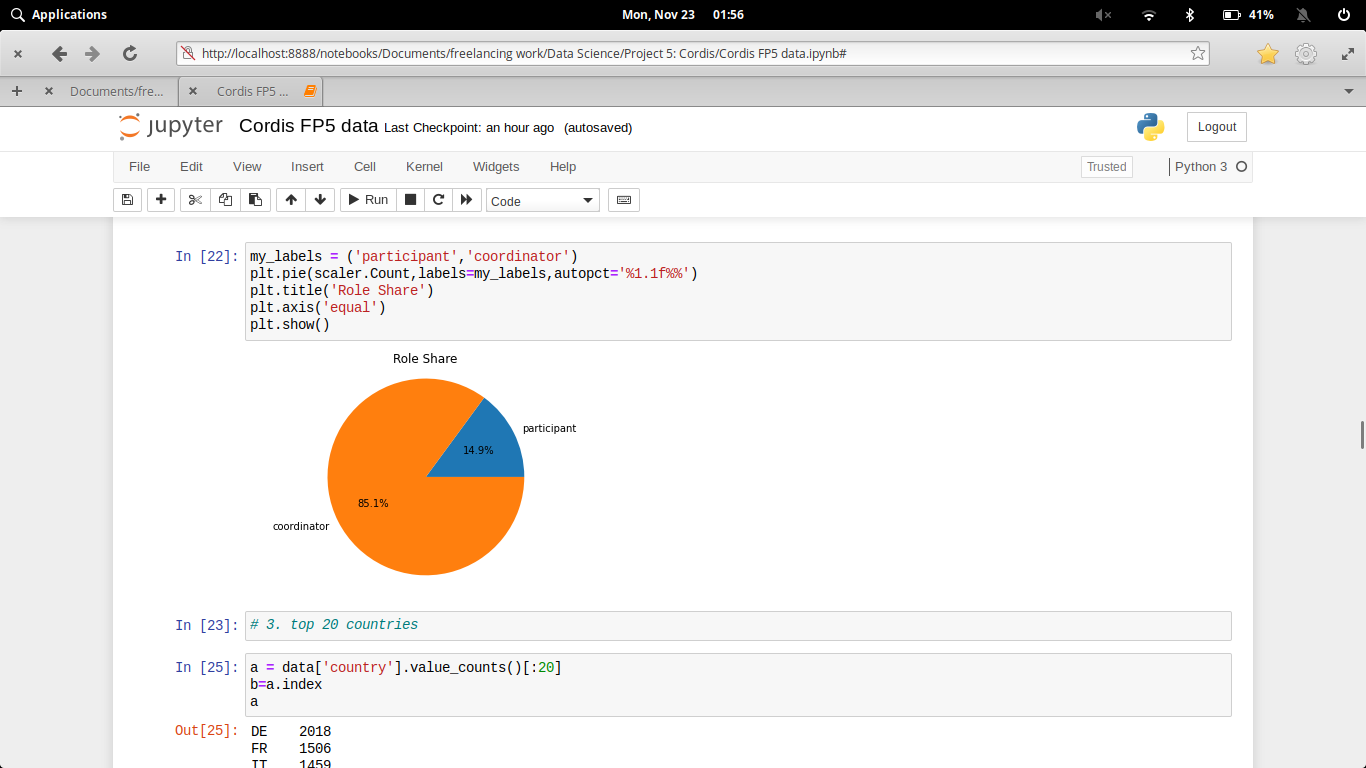


**Output:-**

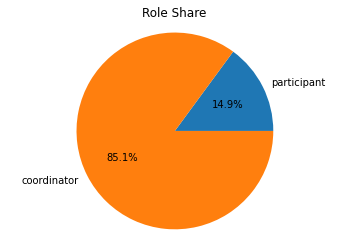
****

1. **What is the coordinator to participant ratio?**

**Code:-**



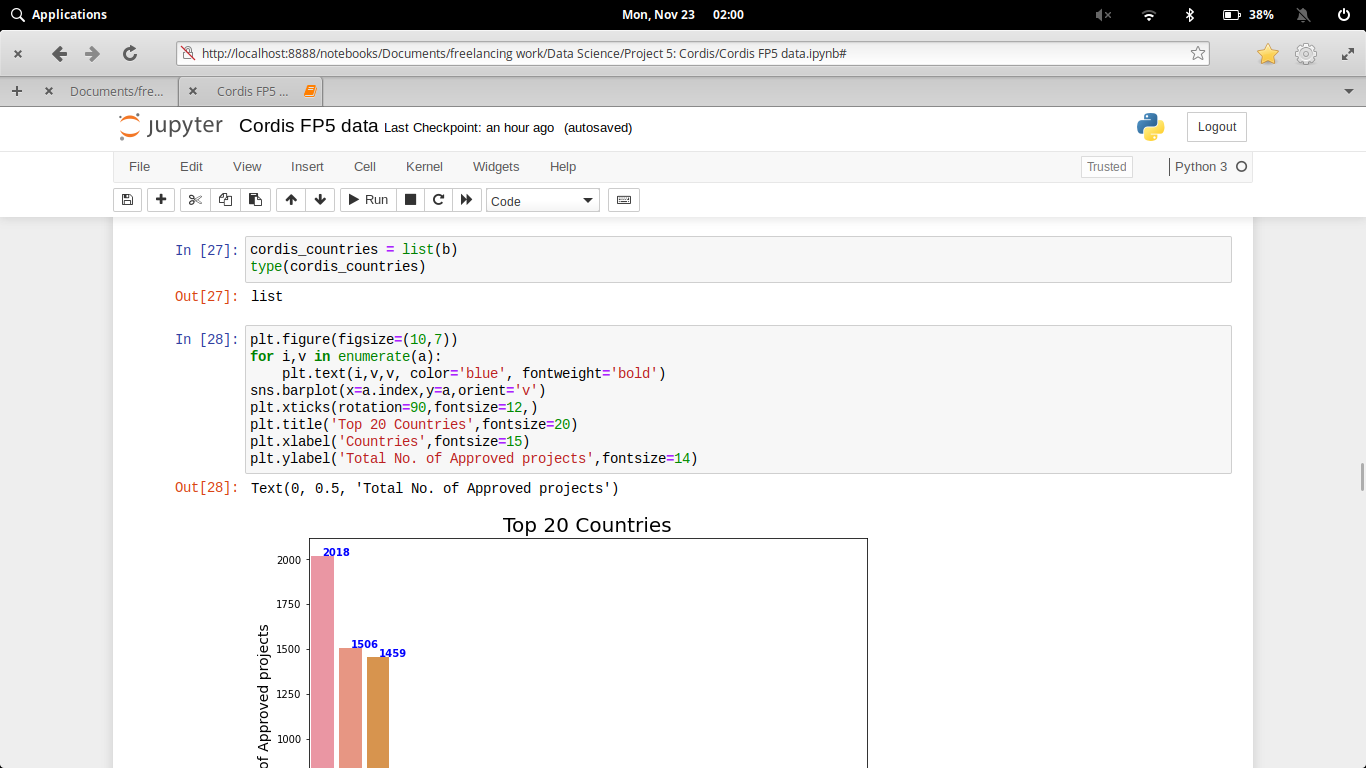
**Output:-**



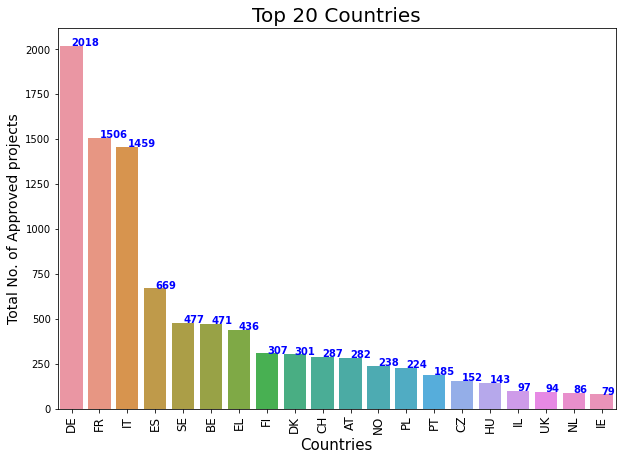
This shows that there are coordinates way more than the number of participants.

1. **Which are the top 20 countries with maximum investment?**

**Code:-**

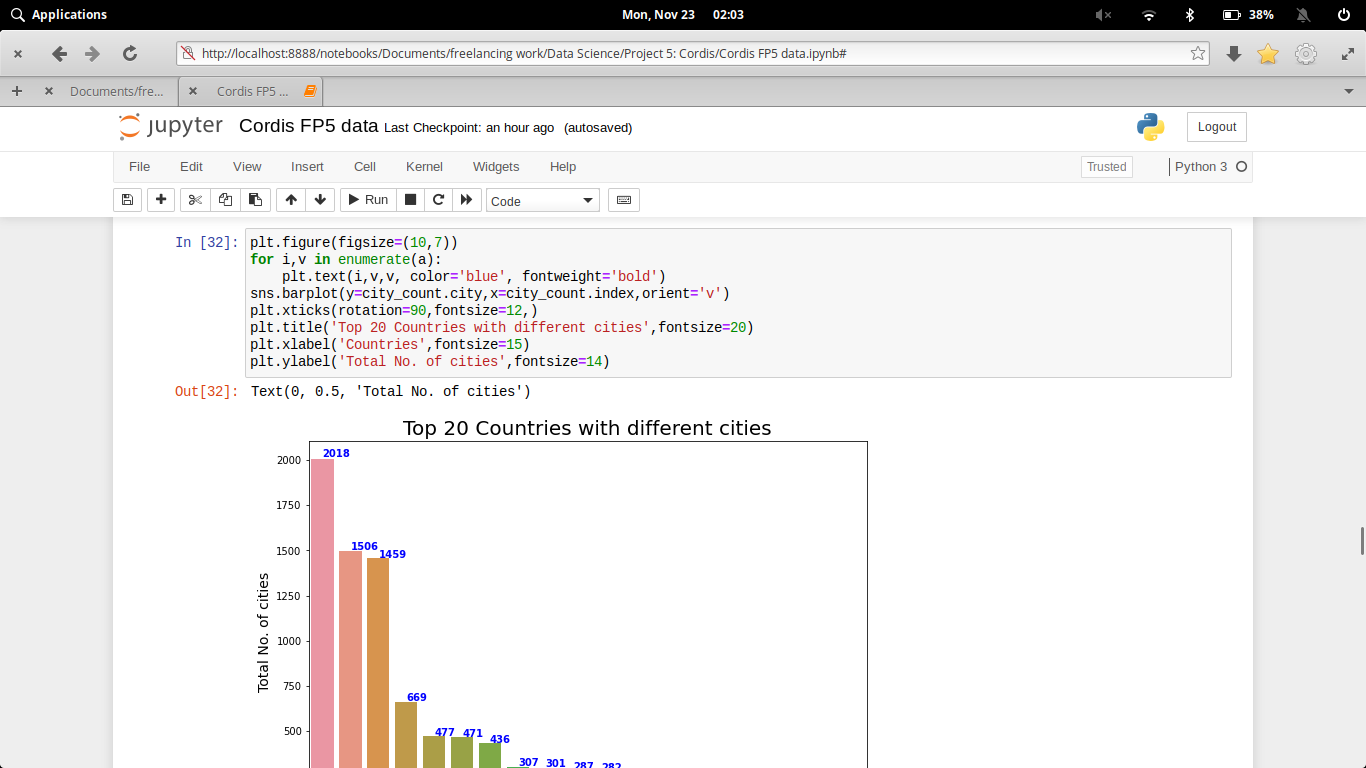


**Output:-**

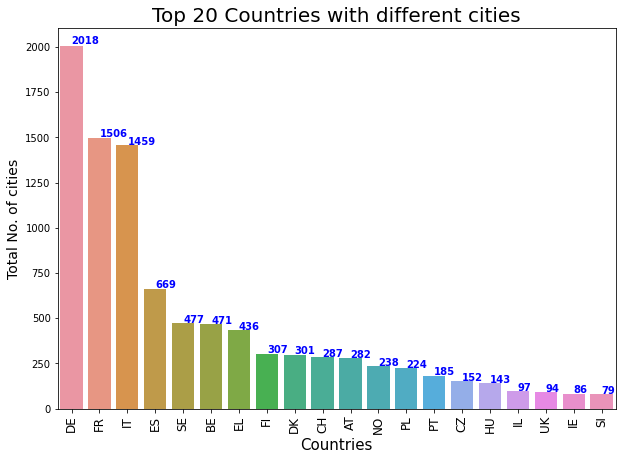


1. **Which top 20 countries contributed from maximum number of states?**

**Code:-**

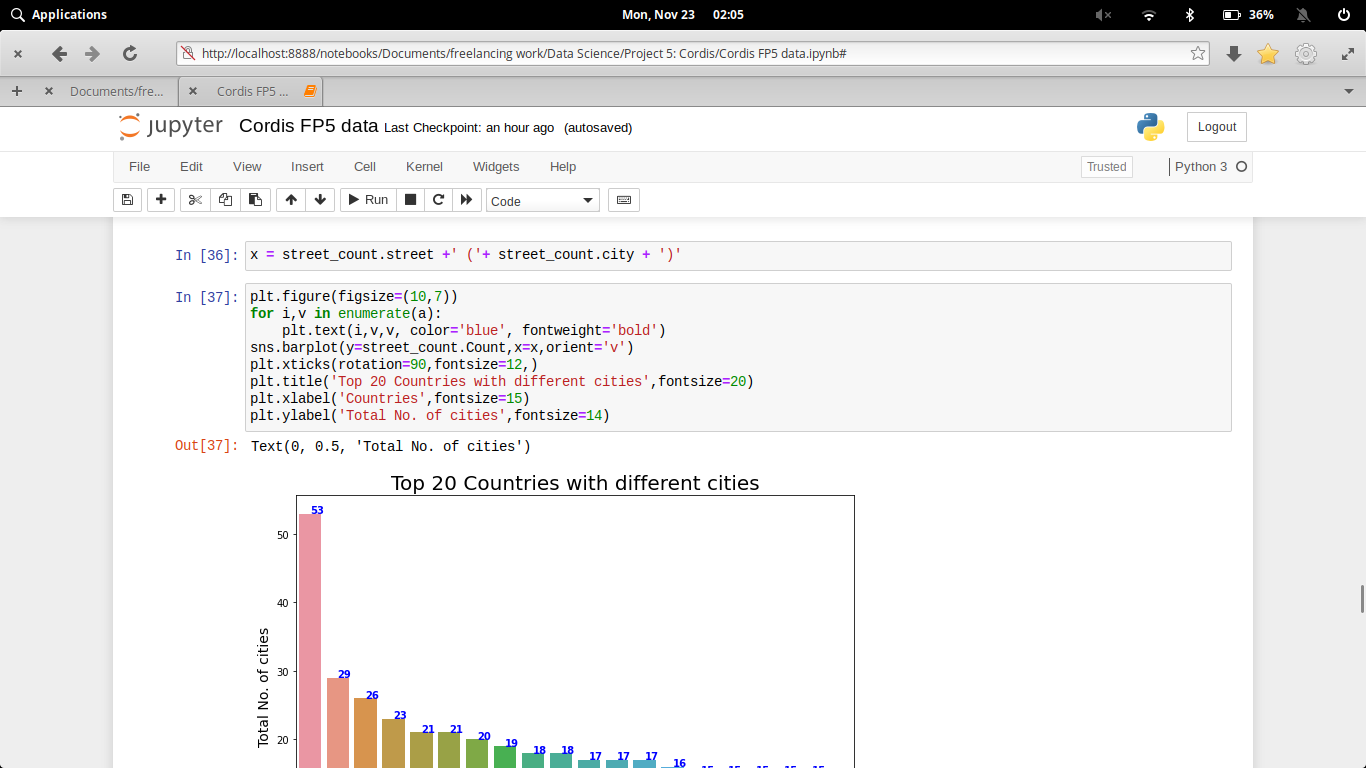


**Output:-**

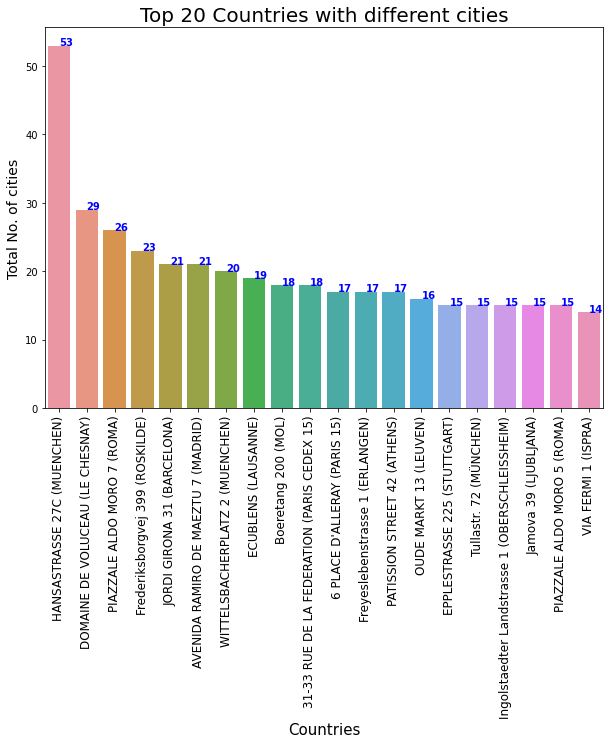


1. **Which locality(street) is the major contributor in the cordis growth?**

**Code:-**

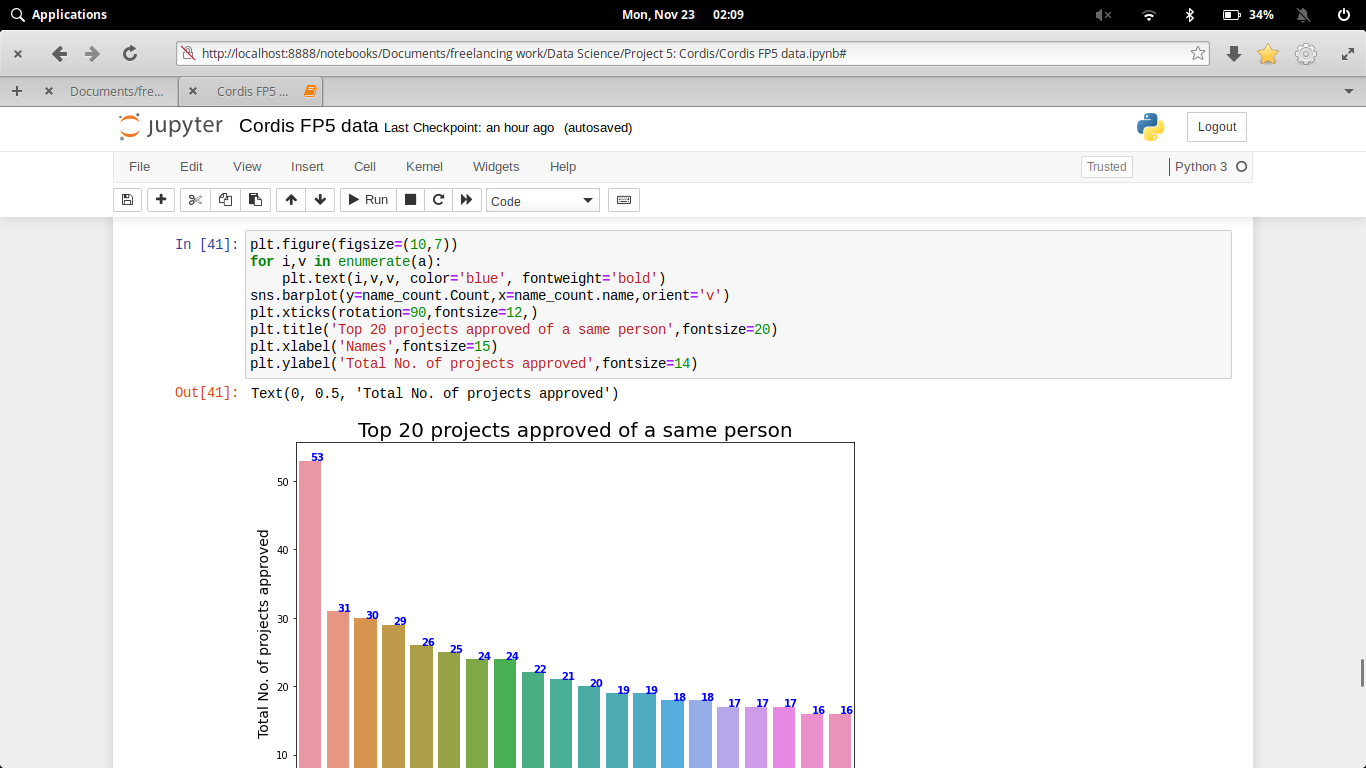
****

**Output:-**

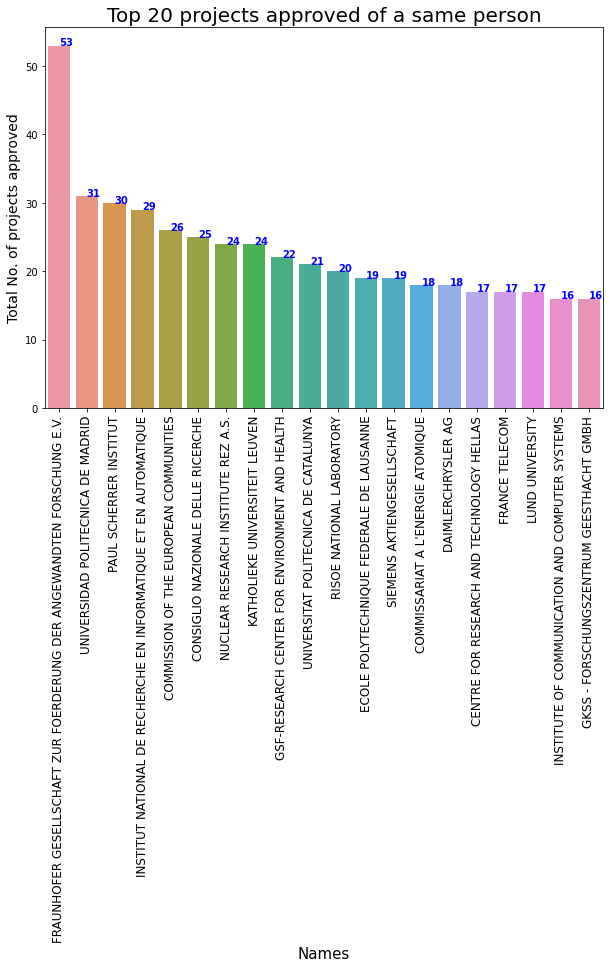
****

1. **What is the max number of projects approved with a single identity?**

**Code:-**

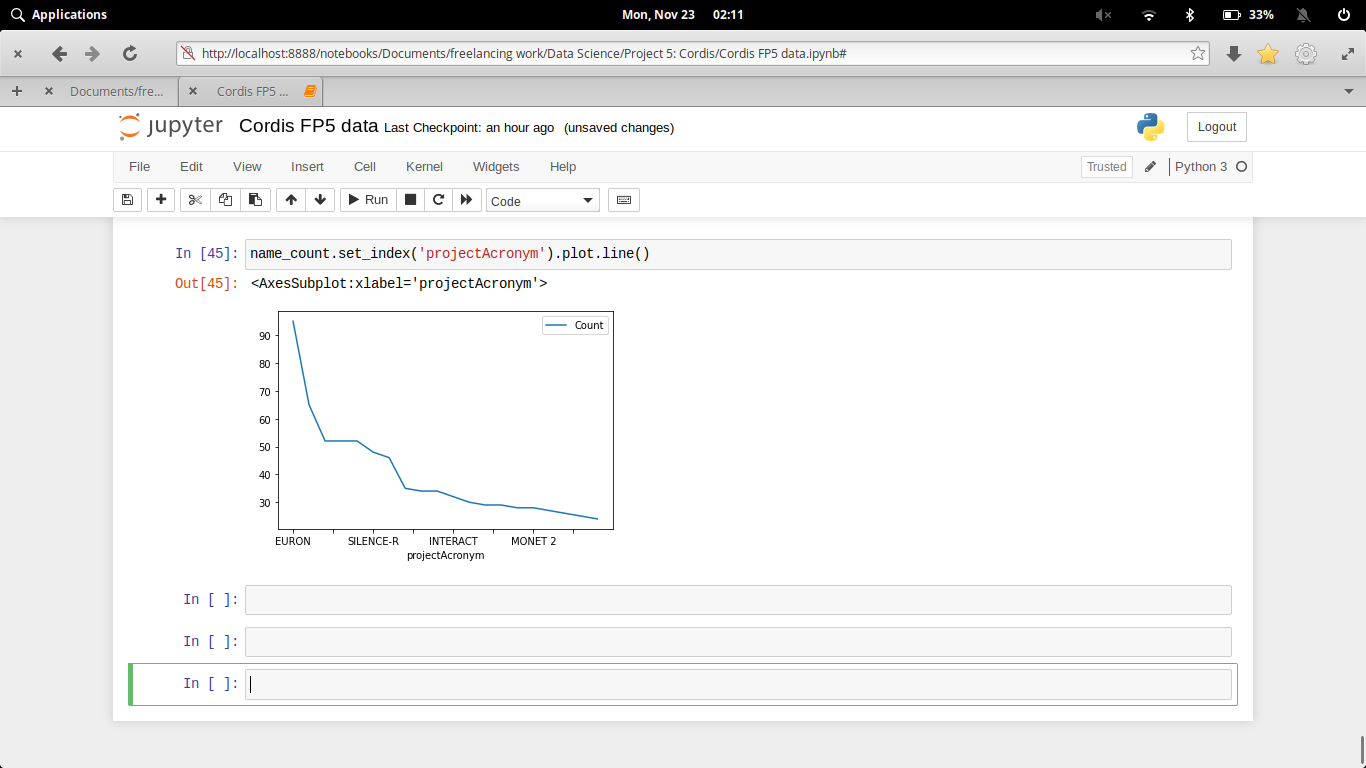


**Output:-**

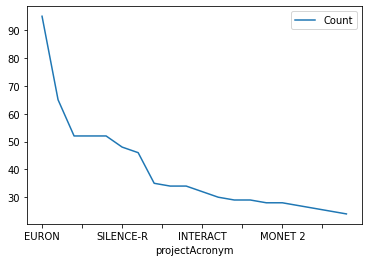
****

1. **What are the some top projectAcronym approved by FP5**

**Code:-**



**Output:-**



# Executive Summary

**CONCL****USION**

From this above analysis, we conclude that:

1. There are participants way less than the coordinators.
2. Germany, followed by France and Italy have highest number of contributors with 2018, 1506, and 1459 respectively.
3. Projects like EURON, EUTIST-IMV, EUTIST-AMI are have maximum number of research going on.

# References

* https://cordis.europa.eu/project/id/7210-PR-313
* <https://data.europa.eu/euodp/en/data>
* https://en.wikipedia.org/wiki/Cordis\_(medical)