**INTRODUCTION:**

* This dataset content information on startup name, industry vertical, location, investors of the startup etc.
* We try to understand the dataset by answering questions like which startup is funded the most, which is the preferred industry vertical for investor, which city are startup preferring etc.

**ABSTRACT:**

The craze among Indian Graduates or student still studying in Various Engineering colleges and Business schools to build their own company right from scratch has been at a peak in the last decades. Although, as a country we have seen examples of various companies starting from small firms to Huge MNC’s like Infosys, Reliance etc. so for proper understanding about which startup will be in boom in coming area for it we are going to see the Data analysis process and kind of conclusion that can be drawn from ‘*Indian Startup Funding Dataset*’ .

**Here we will go through an agenda of some proper questions about to help in exploring Indian Startup Funding Dataset.**

**Questions:**

1. How much funds does startups generally get in India? (Maximum and Minimum funding).
2. Which industry are favored by investors for funding? Or which type of companies got more easily funding?
3. Do cities play a major role in funding? Or which city has maximum startups?
4. Who is the important investor in the Indian Ecosystem?
5. What are different types of funding for startups?

**Data Wrangling/Munging the Data...!!**

Data wrangling sometimes referred to as data munging is the process of transforming and mapping data from one “raw” data form into more appropriate and valuable for a variety of downstream purposes such as analytics.

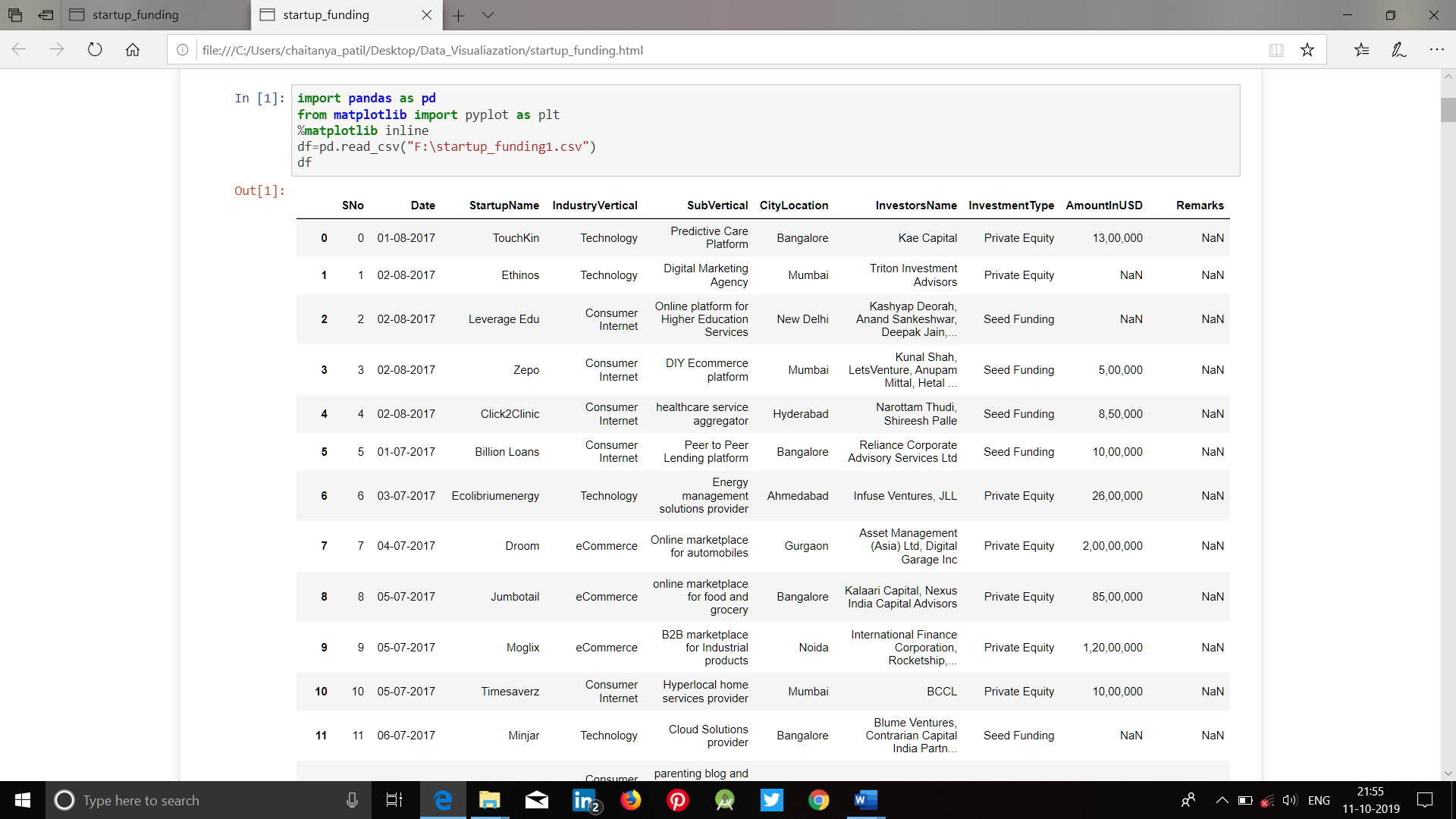
So in this step we are going to access, clean our dataset for more appropriate use during our further analysis process.

**Data Wrangling has 3 steps:**

1. **Gathering Data:**

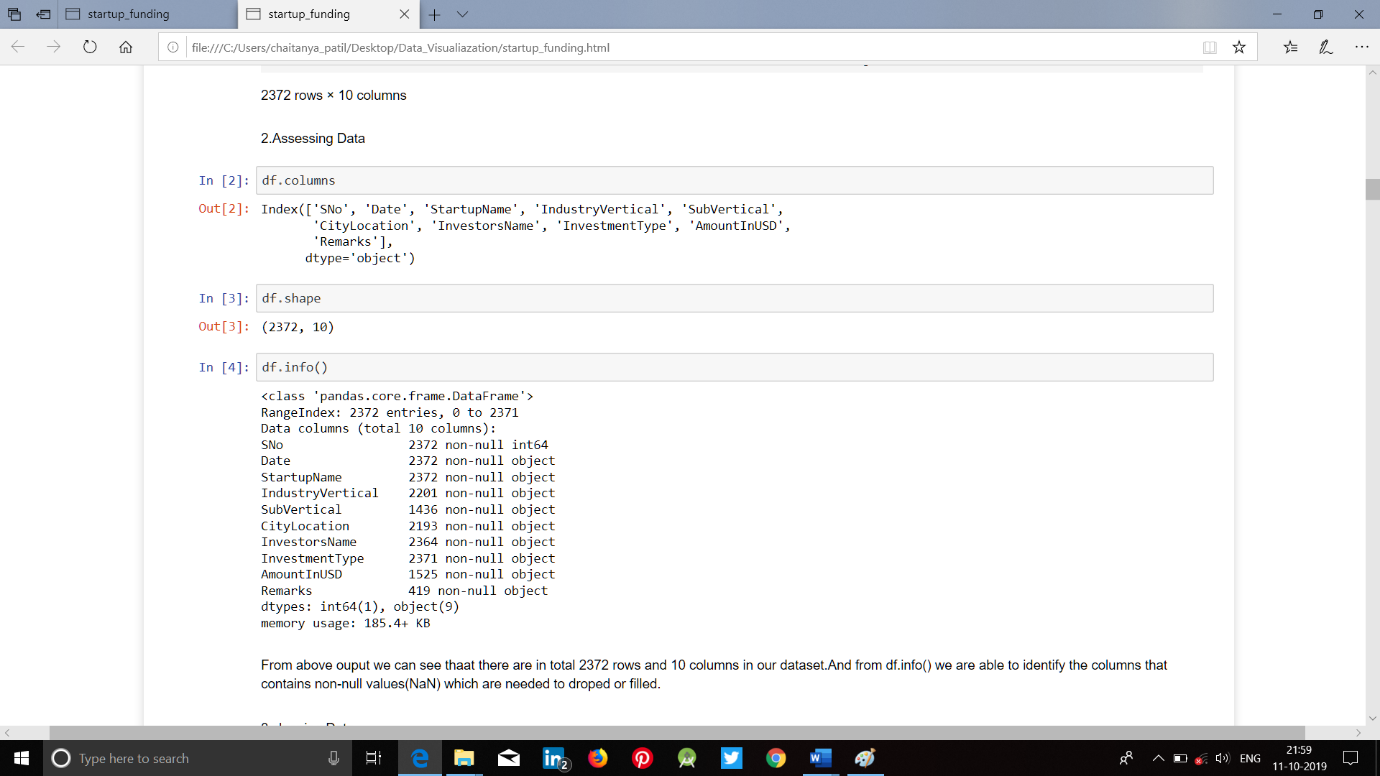
In this we gather data from various sources like CSV files, API, Databases or Web Scraping.

The data of *Indian Startup Funding* is in CSV format, so we just have to import the libraries and read our CSV file, And our dataset is ready for the analysis process.



1. **Assessing Data:**

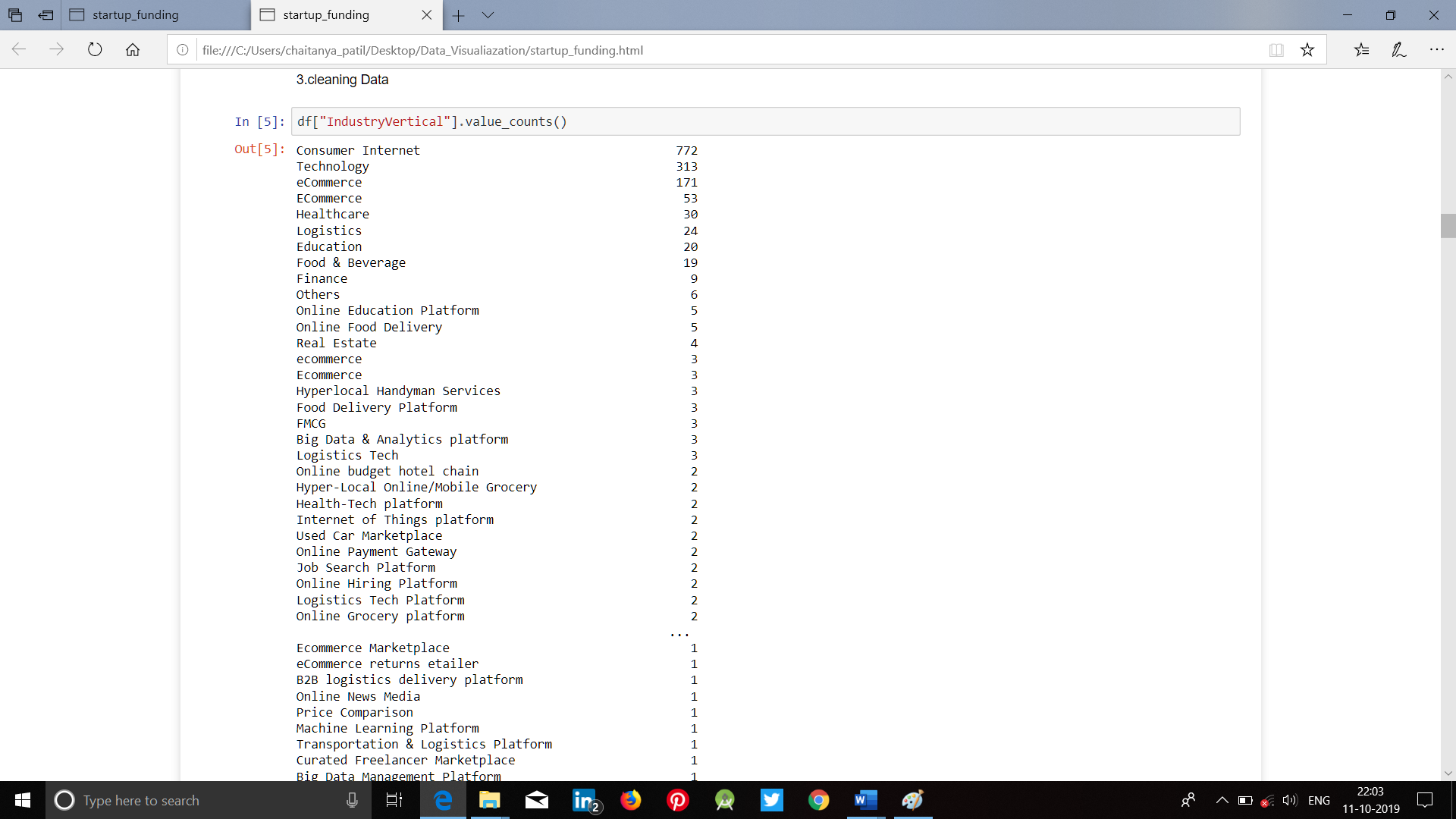
After gathering the data now we have to do some analysis to get the proper idea about out dataset-like how many *rows and colums* are present in our dataset,or what is the *shape* of our data,or finding the *info* about our data.



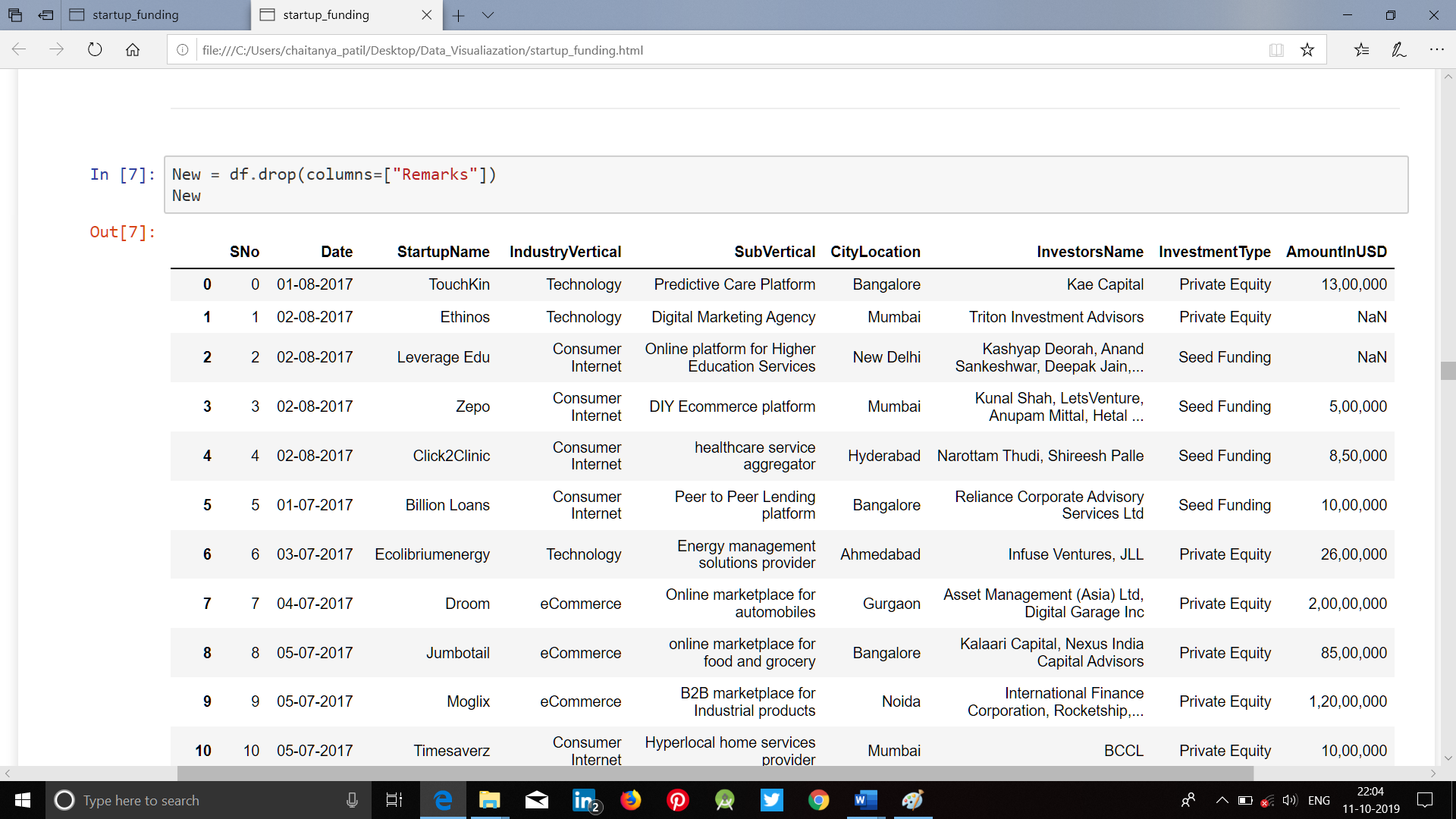
3. **Cleaning Data:**

Here,we do the analysis for detecting and correcting missing, or inaccurate records from a data set.

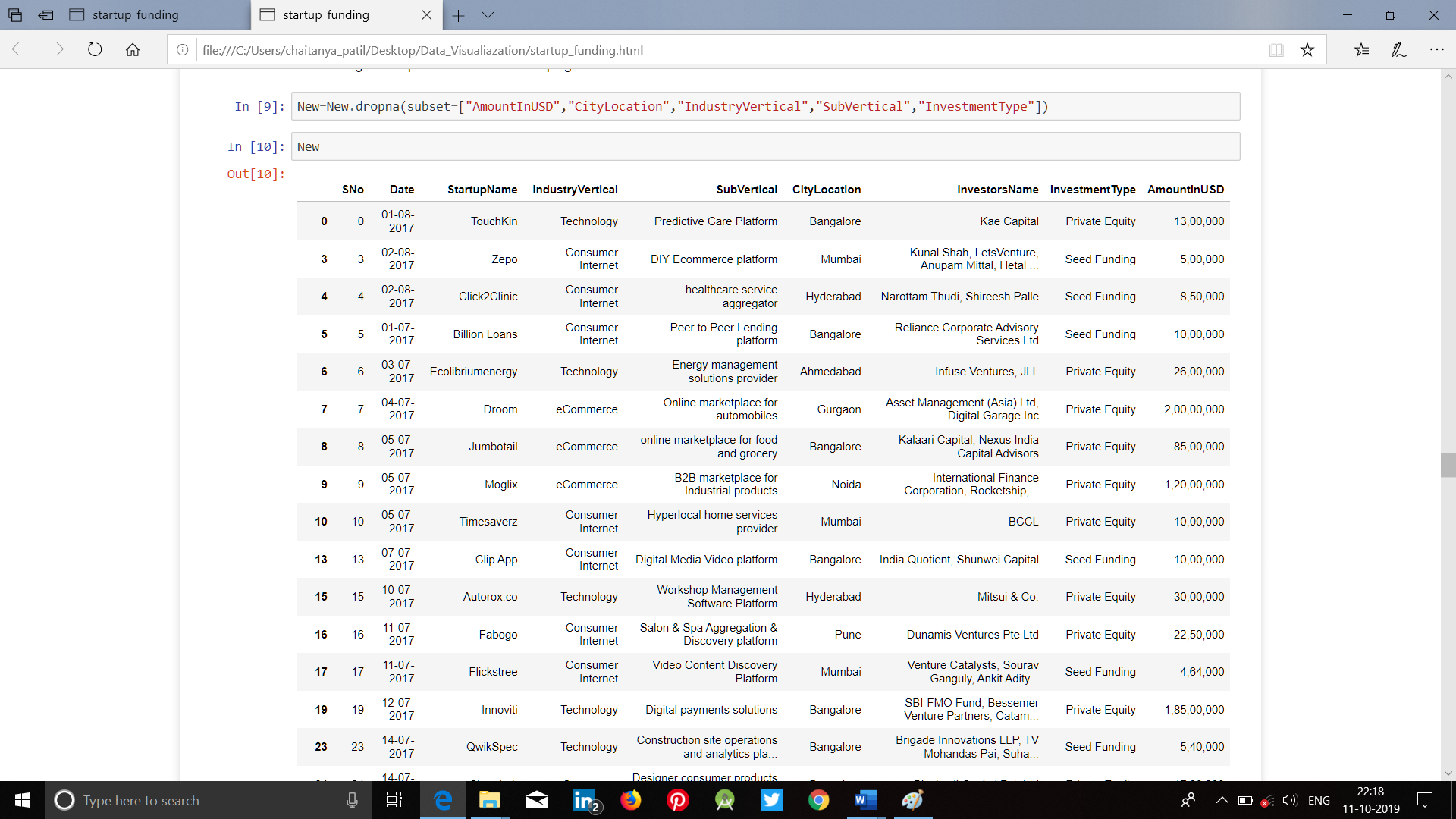
*As we are doing our analysis on Indian Startup Funding Dataset so lets see some of the data cleaning process that we need to do in this-*



From above output we can see that the value counts on IndustryVertical column,here almost 80% of the data or values we are getting from top three fields i.e.,***Consumer interest, Technology & eCommerce***.So to make our data more appropriate for analysis we are only taking these three category by using ***isin*** function and rest of the data of IndustryVertical column we are discarding them.



Here we are using the ***dropna*** function in columns like ***AmountInUSD, CityLocation, IndustryVertical, InvestmentType*** to drop the non-null values.



**STEP 3: Data Analysis..!!**

Once the data is collected, cleaned, and processed, it is ready for Analysis.Here, we have to find the correlation, covariance of the data using graph,barplot,pieplot etc.

**STEP 4: Drawing Conclusions..!!**

Now as we have done our analysis,we are able to draw conclusion or we can make some prediction.

From this dataset we can draw some conclusion based Descriptive analysis such as-

1. Bangalore has the highest number of startup.
2. Most percent of Investment type is private equity.
3. Consumer Internet Industry uses most seed funding investment.
4. Maximum funding that startups generally get in India is about ***971000.***
5. Minimum funding that startups generally get in India is about ***100000.***