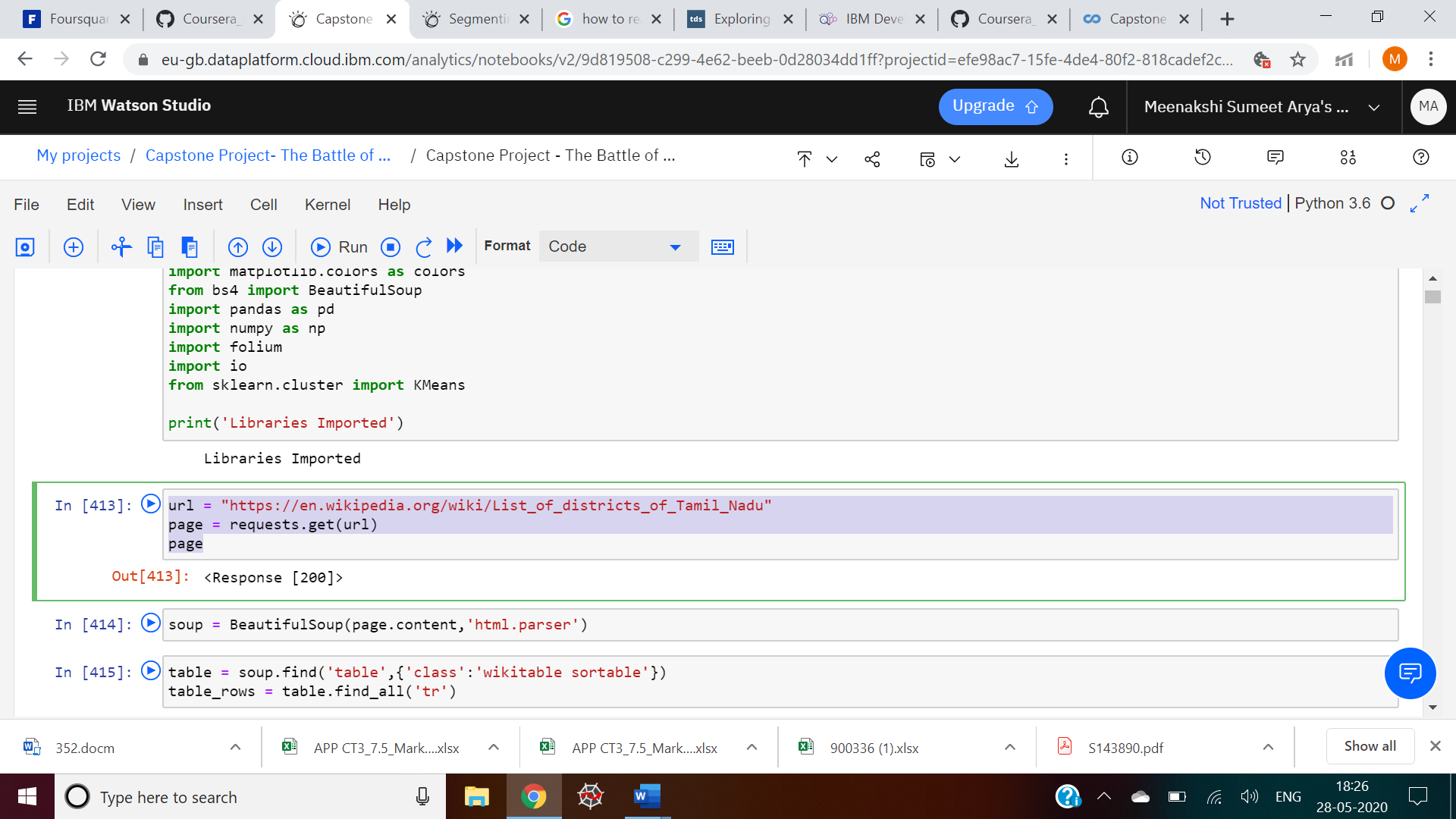
**Data Description**

**Prospects Of Starting a Multiplex In The Most Densely Populated District of Tamilnadu, India**

Data Source: <https://en.wikipedia.org/wiki/List_of_districts_of_Tamil_Nadu> The dataset at the given link will be used for obtaining the relevant information about the various districts in Tamilnadu, their area, population and talukas within each district. This data alongwith Foursquare location data will be used for obtaining the information about the number of multiplexes in the most populous districts of Tamilnadu.

Data Preparation:

Scraping Tamilnadu districts from Wikipedia was done using libraries “requests” and “BeautifulSoup(imported from bs4)” the table was extracted containing the un-processed data.

The table needs a lot of preprocessing:

1. All rows with more than 2 Null values will be removed.
2. The columns with values separated by “\n” will be separated by “,”.
3. Certain columns which are not required for our specific problem definition will be dropped.

Post preprocessing, the data frame looks like this:

