**EDA & Stats Project**

|  |  |  |
| --- | --- | --- |
| **Topics** | **Remarks** | **Max Marks** |
| EDA & Stats | **Solve the Problems given below in the Task** | **100** |

**Domain:**

Customer Service data set

**About:**

NYC Open Data is a vast trove of City government datasets that have been made available to the public. One such dataset, 311 Service Requests from 2010 to Present, will be the focus of this article. These 311 data contains information about more than 24 million service requests made since 2010. For those who aren’t familiar, 311 is a phone number used in the U.S. that allows callers to access non-emergency municipal services, report problems to government agencies, and request information.

This article discusses my process for exploring trends in a recent subset of the data (June-Nov. 2020), and for building a neural network to classify the government agency that responded to a given call. The NYPD is responsible for responding to most 311 calls, and some agencies respond to very few.

**What is Expected?**

Being a data analyst, you must Perform a service request data analysis of New York City 311 calls. You will focus on the data wrangling techniques to understand the pattern in the data and also visualize the major complaint types. You are also expected to do some basic descriptive analysis that you think highlights important outcomes/findings from the data.

**Task - Data Analysis using Python.**

(Perform a service request data analysis of New York City 311 calls)

* Import a 311 NYC service request. Do data cleaning (Ex – drop empty columns, drop unnecessary columns and so on). **(20 marks)**
* Drop rows where closed and resolution date both are empty and update closed date with resolution date where closed date is empty and vice versa. **(10 marks)**
* Read or convert the columns ‘Created Date’ and Closed Date’ to datetime datatype and create a new column ‘Request\_Closing\_Time’ as the time elapsed between request creation and request closing. (Hint: Explore the package/module datetime). **(10 marks)**
* Provide major insights/patterns that you can offer in a visual format (graphs or tables); at least 4 major conclusions that you can come up with after generic data mining. **(20 marks)**
* Order the complaint types based on the average ‘Request\_Closing\_Time’, grouping them for different locations. **(10 marks)**
* Perform a statistical test for the following: **(30 marks)**
* Whether the average response time across complaint types is similar or not (overall)
* Are the type of complaint or service requested and location related?

**Please note:** For the above statements you need to state the Null and Alternate and then provide a statistical test to accept or reject the Null Hypothesis along with the corresponding ‘p-value’.