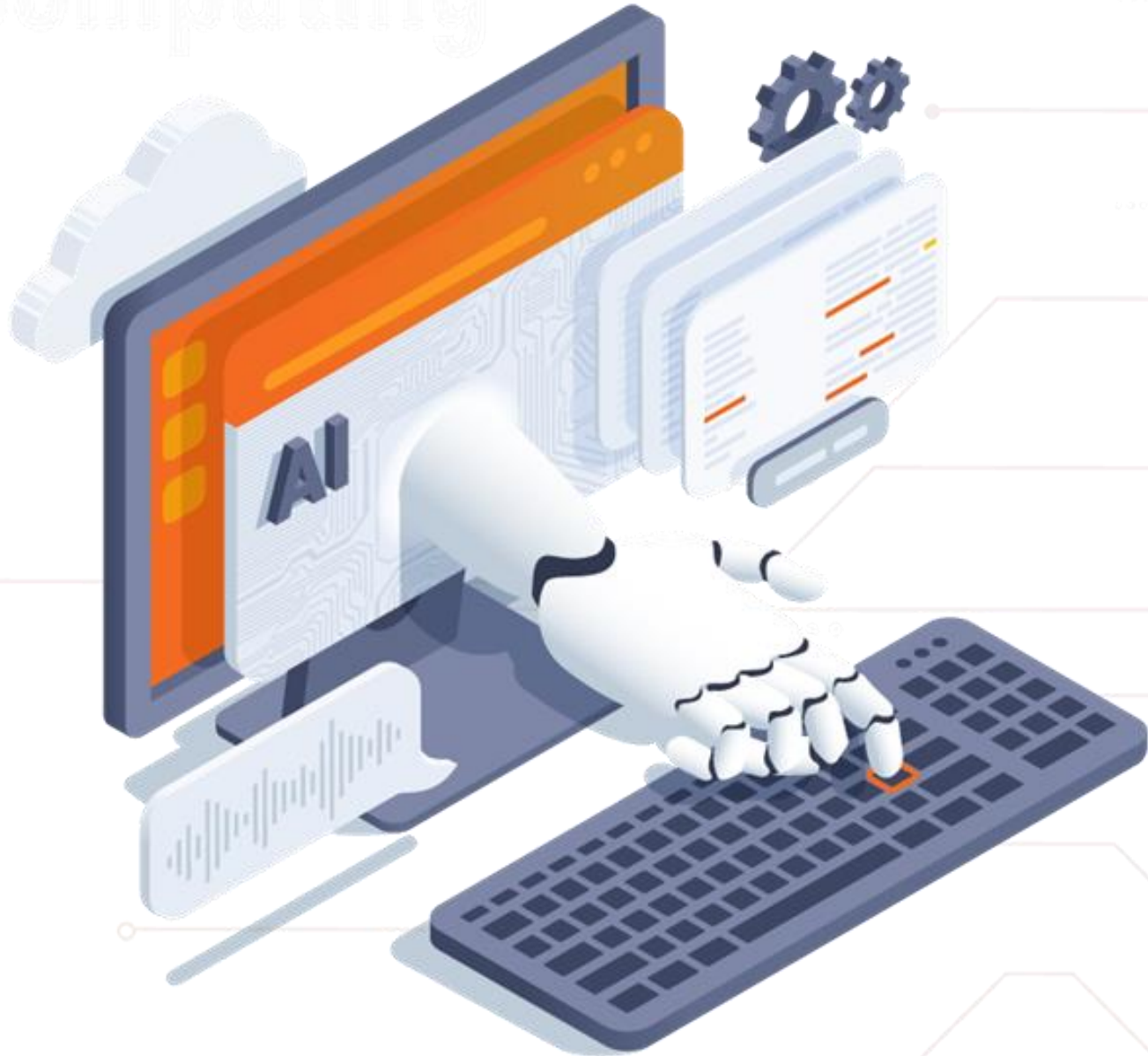


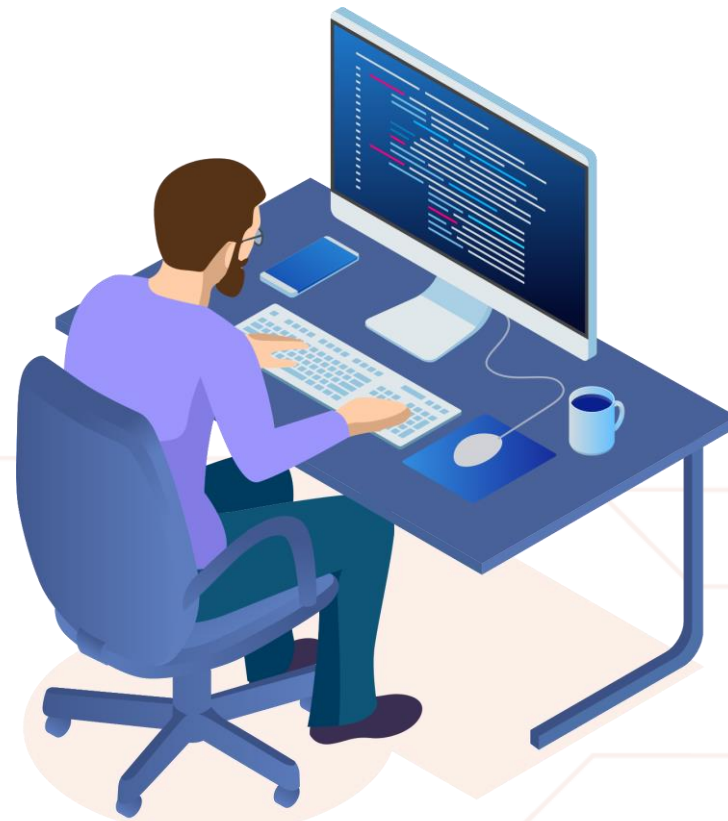
Cloud
Computing



Caltech

**Center for Technology &
Management Education**

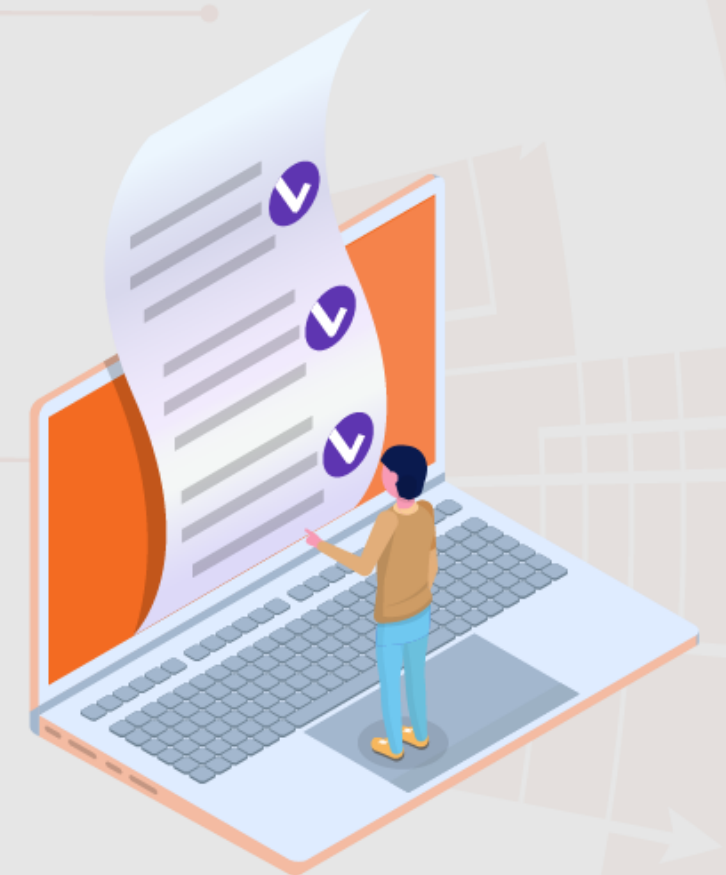
Applied Data Science with Python Course-end Project



Customer Service Requests Analysis

Objectives

- To assess the data and prepare a fresh dataset for training and prediction
- To plot a bar graph to identify the relationship between two variables
- To visualize the major types of complaints in each city



Prerequisites



- Basics of Python
- Application of Python libraries in data science
- Perform analysis on a dataset
- Knowledge of DataFrame
- Train and perform prediction on a dataset

Industry Relevance



- **Basics of Python:** It is used for web development, data science, data analysis, and machine learning and in startups and the finance industry.
- **Application of Python libraries in data science:** Python's large library ecosystem makes it possible to perform a wide range of functions, particularly in data science and machine learning.
- **Perform analysis on a dataset:** Data analysis refers to the process of manipulating raw data to uncover useful insights and draw conclusions.

Industry Relevance



- **Knowledge of DataFrame:** DataFrames are one of the most common data structures used in modern data analytics as they are a flexible and intuitive way of storing and working with data.
- **Train and perform prediction on a dataset:** The initial dataset used to train machine learning algorithms is known as training data. This data is used by models to develop and improve their rules.

Problem Statement



You've been asked to analyze data on service request (311) calls from New York City. You've also been asked to utilize data wrangling techniques to understand the patterns in the data and visualize the major types of complaints.

Note: Download the **311-service-requests-nyc.zip** file using the link given in the **Customer Service Requests Analysis** project problem statement and extract the **311_Service_Requests_from_2010_to_Present.csv** file

Dataset Description

Dataset name: 311-service-requests-nyc.csv



Variable	Description
Unique Key	The unique identification number
Created Date	The date when the request was created
Closed Date	The date when the request was closed
Agency	The agency that handled the case
Agency Name	The full name of the agency that handled the case
Complaint Type	The type of complaint received
Descriptor	The description of the complaint
Location Type	The type of location where the incident occurred
Incident Zip	The zip code of the location
Incident Address	The location at which the incident occurred

Dataset Description

Dataset name: 311-service-requests-nyc.csv



Variable	Description
Street Name	The name of the street
Cross Street 1	The cross of the street 1
Cross Street 2	The cross of street 2
Intersection Street 1	The first point of intersection of both streets
Intersection Street 2	The second point of intersection of both streets
Address Type	The type of the address
City	The city where the incident occurred
Landmark	The landmark near the incident that occurred
Facility Type	The type of the facility
Status	The status of the complaint

Dataset Description

Dataset name: 311-service-requests-nyc.csv



Variable	Description
Due Date	The due date of the complaint
Resolution Description	The resolution provided by the police department
Resolution Action Updated Date	The date at which the resolution was provided
Community Board	The location of the community board
Borough	The town, or area inside a large town, that has some form of local government
X Coordinate (State Plane)	The X coordinate of the plane
Y Coordinate (State Plane)	The Y coordinate of the plane
Park Facility Name	The name of the park facility
Park Borough	The park town, or area inside a large town, that has some form of local government
School Name	The name of the school (optional)

Dataset Description

Dataset name: 311-service-requests-nyc.csv



Variable	Description
School Number	Number of the school (optional)
School Region	Region of the school (optional)
School Code	Code of the school (optional)
School Phone Number	Contact information of the school (optional)
School Address	Address of the school (optional)
School City	City at which the school is located (optional)
School State	State in which the school is located (optional)
School Zip	Zip code of the school (optional)
School Not Found	Valid if the school is not found (optional)
School or Citywide Complaint	Contains the complaint of the school (optional)

Dataset Description

Dataset name: 311-service-requests-nyc.csv



Variable	Description
Vehicle Type	Type of vehicle used (optional)
Taxi Company Borough	Information on the taxi company (optional)
Taxi Pick Up Location	Pick up location of the taxi (optional)
Bridge Highway Name	Name of the highway bridge (optional)
Bridge Highway Direction	Direction of the highway bridge (optional)
Road Ramp	Information on the road ramp (optional)
Bridge Highway Segment	Segment of the bridge (optional)
Garage Lot Name	Name of the garage (optional)
Ferry Direction	Ferry direction information (optional)
Ferry Terminal Name	Name of the ferry terminal (optional)
Latitude	Latitude value
Longitude	Longitude value
Location	Location information

Tasks to Perform



1. Understand the dataset:
 - 1.1 Import the dataset
 - 1.2 Visualize the dataset
 - 1.3 Print the columns of the DataFrame
 - 1.4 Identify the shape of the dataset
 - 1.5 Identify the variables with null values
2. Perform basic data exploratory analysis:
 - 2.1 Draw a frequency plot to show the number of null values in each column of the DataFrame
 - 2.2 Missing value treatment
 - 2.2.1 Remove the records whose **Closed Date** values are null

Tasks to Perform



- 2.3 Analyze the date column, and remove entries that have an incorrect timeline
 - 2.3.1 Calculate the time elapsed in closed and creation date
 - 2.3.2 Convert the calculated date to seconds to get a better representation
 - 2.3.3 View the descriptive statistics for the newly created column
 - 2.3.4 Check the number of null values in the **Complaint_Type** and **City** columns
 - 2.3.5 Impute the NA value with **Unknown City**
 - 2.3.6 Draw a frequency plot for the complaints in each city
 - 2.3.7 Create a scatter and hexbin plot of the concentration of complaints across Brooklyn

Tasks to Perform



3. Find major types of complaints:

- 3.1 Plot a bar graph to show the types of complaints
 - 3.2 Check the frequency of various types of complaints for New York City
 - 3.3 Find the top 10 complaint types
 - 3.4 Display the various types of complaints in each city
 - 3.5 Create a DataFrame, **df_new**, which contains cities as columns and complaint types in rows
- ## 4. Visualize the major types of complaints in each city
- 4.1 Draw another chart that shows the types of complaints in each city in a single chart, where different colors show the different types of complaints

Tasks to Perform



- 4.2 Sort the complaint types based on the average **Request_Closing_Time** grouping them for different locations
- 5. See whether the average response time across different complaint types is similar (overall)
 - 5.1 Visualize the average of **Request_Closing_Time**
- 6. Identify the significant variables by performing statistical analysis using p-values
- 7. Perform a Kruskal-Wallis H test
 - 7.1 Fail to reject H_0 : All sample distributions are equal
 - 7.2 Reject H_0 : One or more sample distributions are not equal
- 8. Present your observations

Project Outcome



- The project aims to help you work with the dataset and performing analysis.
- In this project, you will assess the data and prepare a fresh dataset for training and prediction.
- You will plot a bar graph to identify the relationship between two variables.
- You will also visualize the major types of complaints in each city.

Submission Process



1. Complete the project in the Simplilearn Lab
2. Complete each task listed in the problem statement
3. Take screenshots of the results for each question and the corresponding code
4. Save it as a document, and submit it using the assessment tab
5. Tap the **Submit** button (this will present you with three choices)
6. Attach the three files, and then click on **Submit**

Note: Be sure to include the screenshots of the output

Thank you