Dalhousie University

CSCI5408 Data Management and Warehousing Analytics

Assignment Report

# Assignment 4

*Submitted By :*
Deeksha Behara :B00784704
Tushar Gupta :B00782699

*Supervisor:*
Suhaib Qaiser

July 3, 2018

# Contents

# 1 Introduction

The motivation of the assignment is learn data analytics using data analytics dashboard. The task also involves plotting graphs, big datasets into human readable format, using visualizing techniques in representing the data and to learn patterns through the data simulation. : https://github.com/bio33/elastic_search_
)

# 2 Experimental Setup - Task Description

The initial task implementation started off by downloading Tableau for desktop and using it to perform various visualization techniques. to represent the data.

## 2.1 Tools, technologies and libraries used

List of tools,technologies and libraries used for the study are as follows:

1. Tableau Desktop:

   Tableau[1] is a tool which assists in exploring and visualizing the data using different graphs provided by them.

   Capabilities of the tool:

   (a) Performing complex aggregations on the data.
   (b) Visualizing the data using appropriate graphs.
   (c) Allows the users to perform spatial joins on the data
   (d) Allows the users to construct geographical graphs
   (e) Can import data from different sources.

# 3 DataSet

The dataset[2] used for the assignment was extracted from hrm-open datasets.The dataset contains the list of permits issued by the HRM's planning and development team. This data was created by the organization for the purpose of issuing the permits.

## 3.1 Dataset Description:

The dataset contains information like Geo-spatial data for varying timestamps.Individual buildings, measurements,type and the community it belongs is mapped to its corresponding estimates price.

## 3.2 Preparing data for Analytics:

Before performing analysis on the data, the data was preprocessed. This included several steps like removing punctuation marks from the text. Additionally, the data checked for nulls and is replaced with suitable values.

## 3.3 Data Loading:

The data is loading into tableau by opening a new sheet and connecting a to csv by clicking on the "other" section. The data loading is straight forward and simple using Tableau.

# 4 Dashboard

The graphs are drawn for each of the queries mentioned in the tasks.

## 4.1 Query1

Use geo spatial visualization to plot the number of permits per location (X and Y are latitudes and longitudes). The size of the point should represent the count of permits per location.

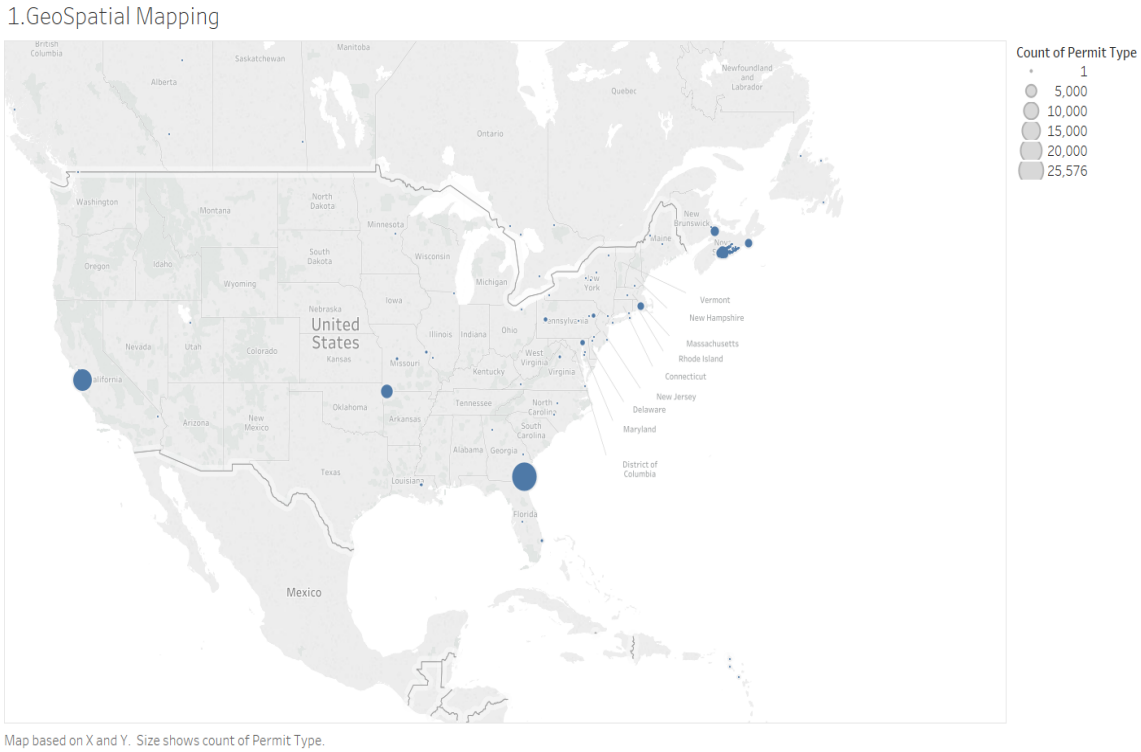


Figure 1: The graph representing the geo spatial representation of the permits per given location represented by the X and Y axes.

## 4.2 Query 2

Create a histogram which represents the count of permits per type (PERMIT_TYPE).

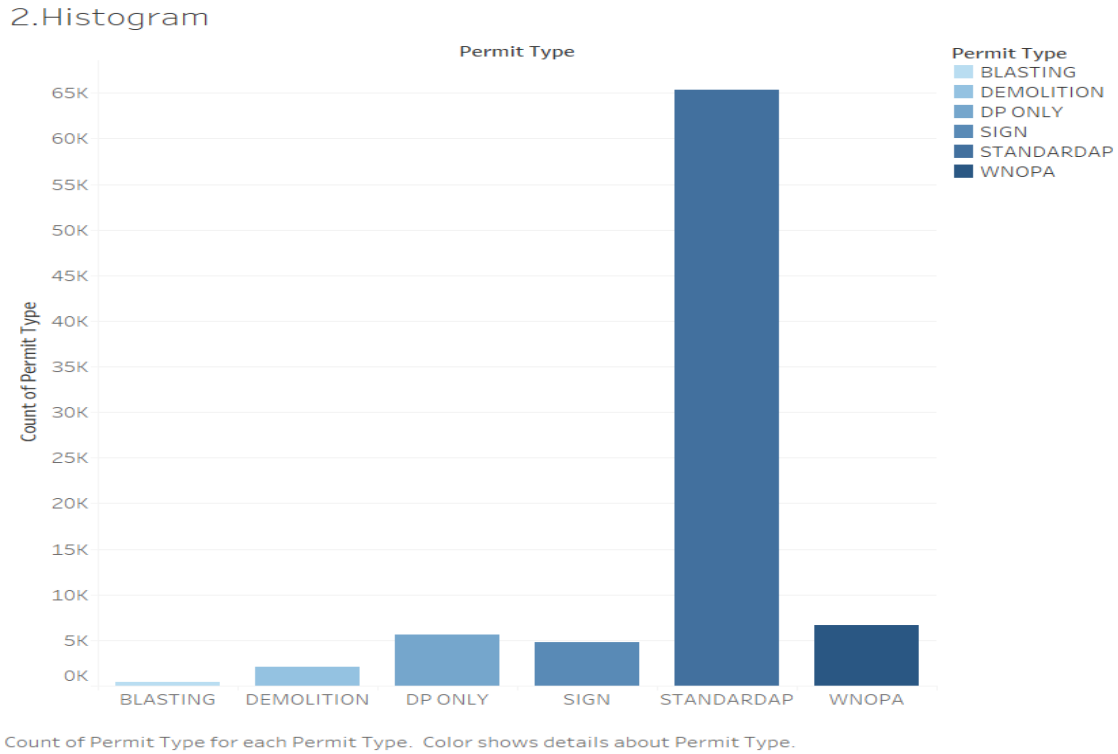


Figure 2: Histograms created for the permits for the given permit_type.

## 4.3 Query 3

Identify time gaps between permit applied and permit issuance date and plot the time delays for all properties. Through visualizing the graph, identify 3-4 properties that took longest time in getting permit.

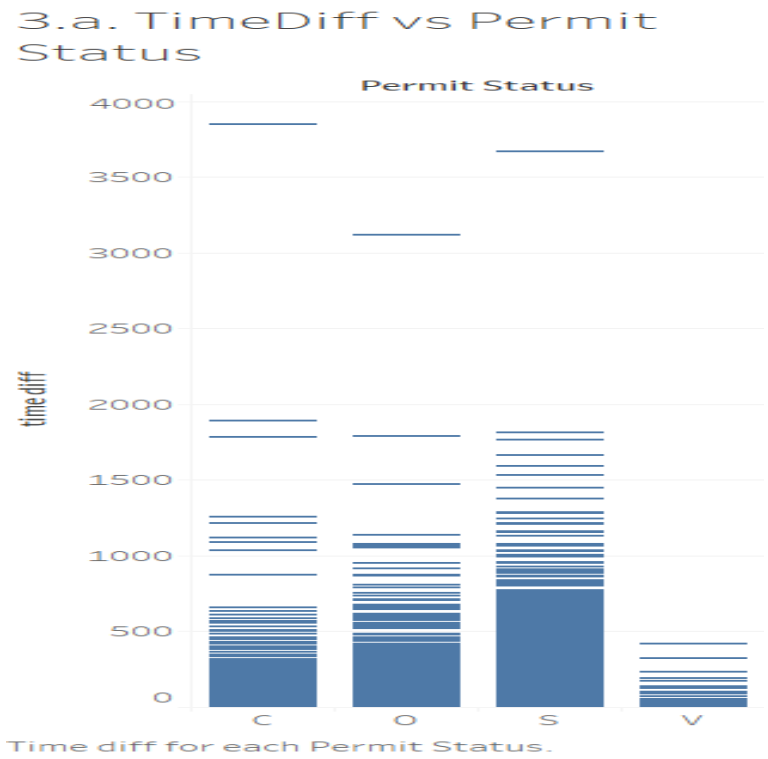


Figure 3: Identifying the time gaps between the permit applied and the issuance date.

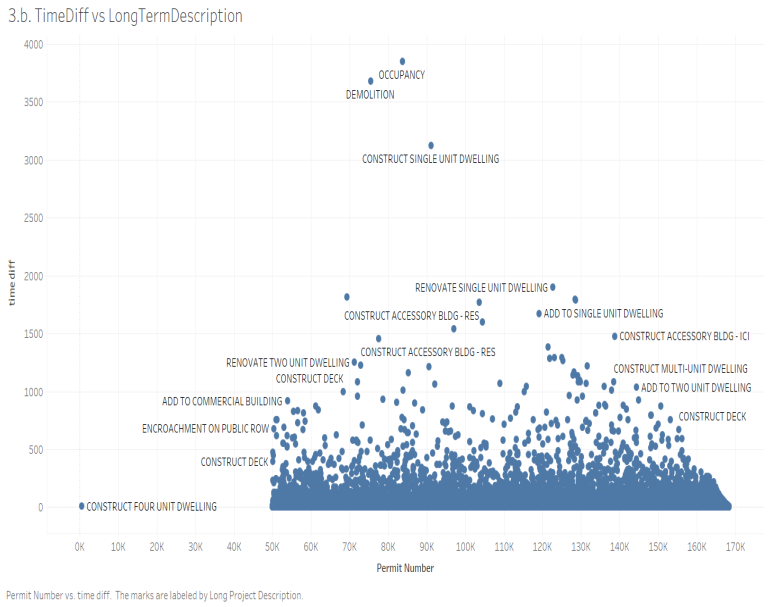


Figure 4: Longest time difference plotted in a graph.

## 4.4 Query 4

Plot estimated value of a project and building type using a plotting mechanism where color coding is applied based on building type.

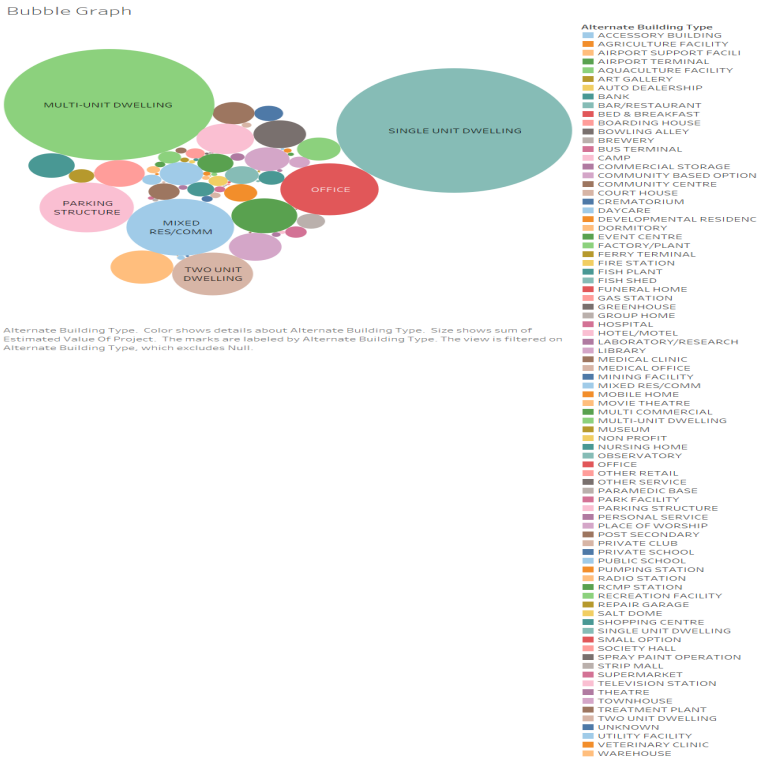


Figure 5: Bubble graph representing the estimated value of the project and the building type.

## 4.5 Query 5

In the above query (4), add another dimension of "Community". Identify which community has the highest number of permits issued for commercial buildings.

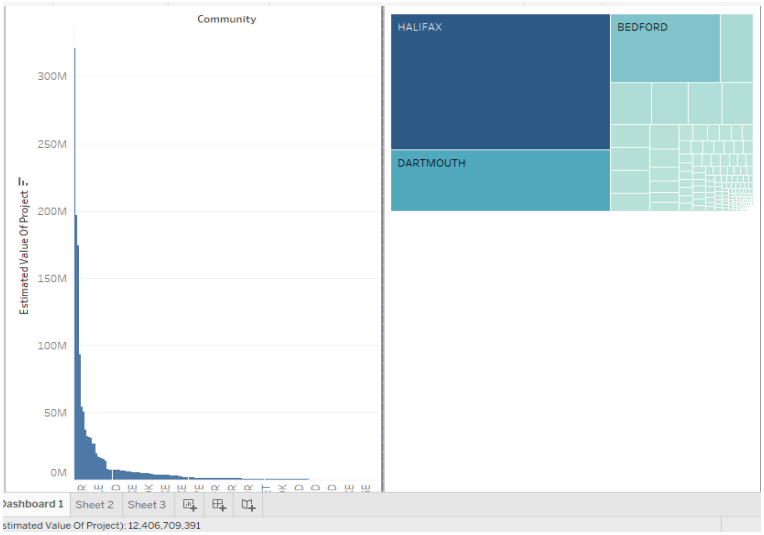


Figure 6: Tree Maps representing the relationship between estimated value, community and the building type.

## 4.6 Output

1. For Query 1, Geo spatial graphs are the most suitable because the mapping involves plotting the data using the latitude and the longitude vales. The size of the Bubbles used for the plotting is directly proportional to the count of permits per location.

2. The query 2 is a simple straight forward graph which involves two variables which are total counts per each permit type.Histograms are the most suitable for this case.

3. Identifying the time difference is depicted using the stacked bar charts and the scattered graphs as shown in Fig 3 and 4. As shown in the graph, the longest time is taken by the permit —

4. Bubble Graphs are a best fit for the Query 4. The size of the bubble represent estimated value of the project. The bubbles are color coded based on the building type.

5. The last graph is represented using the tree maps. The Community which has the highest number of permits is "HALIFAX". However, the results is based on the commercial buildings chosen for the constructing the filter.

# 5 Conclusion

The data was explored and the visual analytics was performed on the data. Tableau was used to represent the data visually using five different graphs.

## References

[1] Tableau. (n.d.).TableauSoftware.(n.d.).Retrievedfromhttps://www.tableau.com/.

[2] Dataset. (n.d.).Retrievedfromhttps://catalogue-hrm.opendata.arcgis.com/datasets/building-permits.