

**CSCI 5408**

**Summer 2018**

**Assignment 4**

**Analyzing data patterns through effective visualization techniques**

**Submitted by:**

**Nikhil Dhirmalani: B00775542**

**Khushboo Siwal: B00781497**

**TABLE OF CONTENTS**

1. Task Description
2. Tool Selection
3. Data Loading
4. Dashboard
5. Output
6. Code Submission
7. References

**LIST OF FIGURES**

[Figure 1 Data Loading from New Data Source 4](#_Toc518309419)

[Figure 2 Connect to the Data file 4](#_Toc518309420)

[Figure 3 Data loaded into Tableau 5](#_Toc518309421)

[Figure 4 Dashboard with all 5 graphs 5](#_Toc518309422)

[Figure 5 Number of permits per location 6](#_Toc518309423)

[Figure 6 Count of permits per type 6](#_Toc518309424)

[Figure 7 Time gaps between permit applied and permit issuance date 7](#_Toc518309425)

[Figure 8 Estimated value of project per building type 8](#_Toc518309426)

[Figure 9 Community with highest number of permits for commercial buildings 9](#_Toc518309427)

1. **TASK DESCRIPTION**

The tasks for this assignment involved us to learn data analytics using any of the data analytical dashboards. The activities involved us to plot graphs by performing some queries on the given data of Building Permits into human readable format.

1. **TOOL SELECTION**

Tableau was chosen for data visualization as it is very fast to deploy and easy to learn. The tool is very intuitive to use. The user interface in Tableau makes it easy to create visual representations of datasets, and the wide range of visualisation types and exploration options gives access to meaningful insights. Tableau also has support for both R and Python. Tableau features better support for connecting to a separate data warehouse.

Tableau has a good speed at building beautiful, graphical analysis and dashboards. The work area in this tool is a straightforward way to make reports and graphs. Tableau has an advantage of working with big data sets. Visuals can be created easily and quickly and can switch between types conveniently to use the model that best represents the task.

1. **DATA LOADING**

Data is downloaded from [https://catalogue-hrm.opendata.arcgis.com/datasets/building-permits](https://catalogue-hrm.opendata.arcgis.com/datasets/building-permits%20location%20in%20csv%20format.) location in csv format. As the data was already cleaned it just had few null rows that we filtered in our tool. After that, New Data source was opened in Tableau and then connecting it to a file saved at a specific location on desktop.

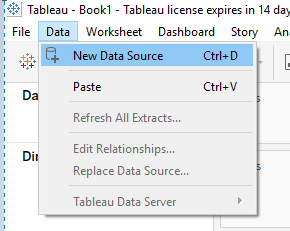


Figure Data Loading from New Data Source

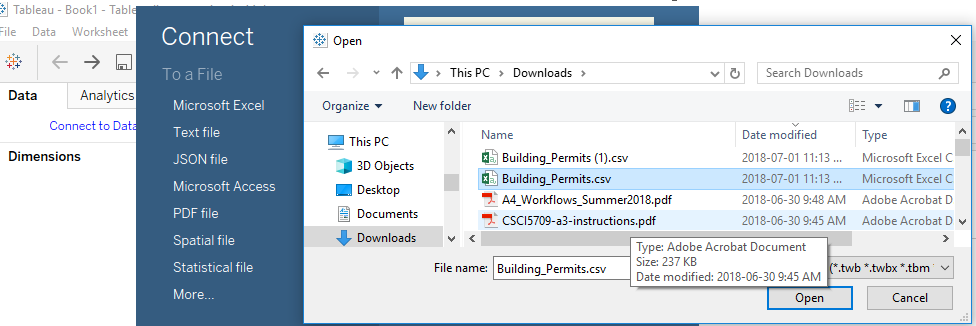


Figure Connect to the Data file

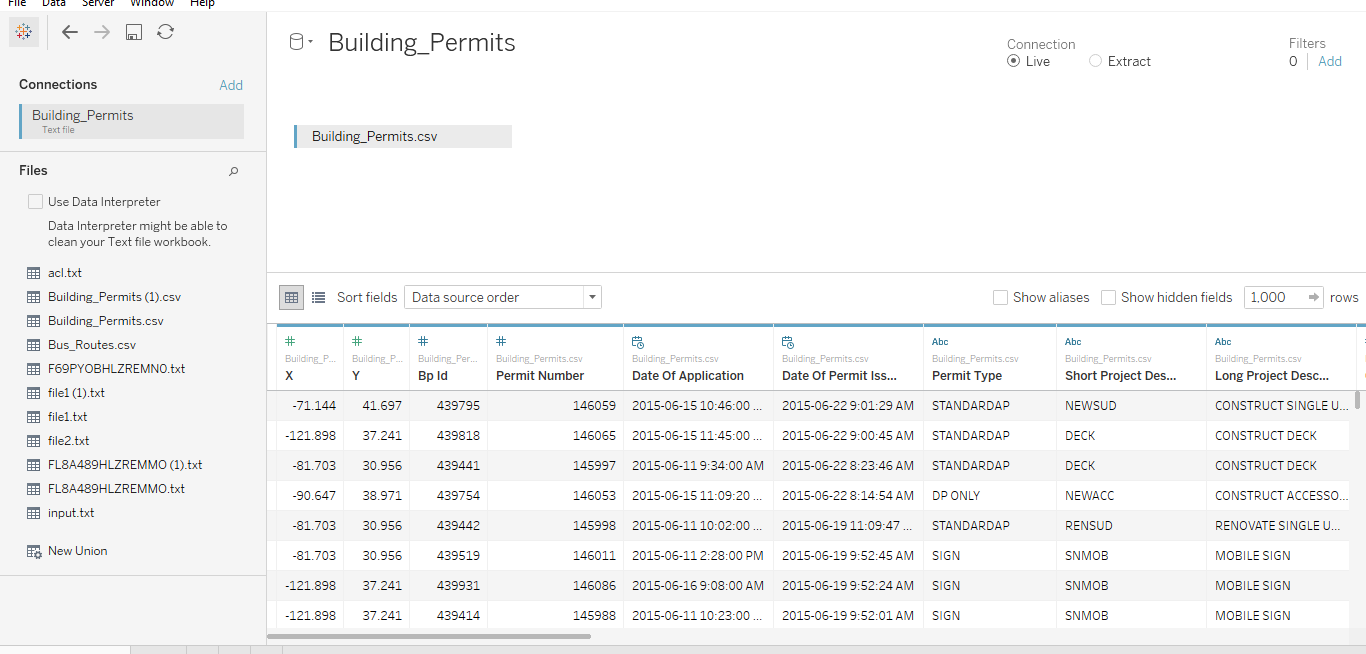


Figure Data loaded into Tableau

1. **DASHBOARD**

Below figure shows the dashboard with all the five graphs used for plotting.

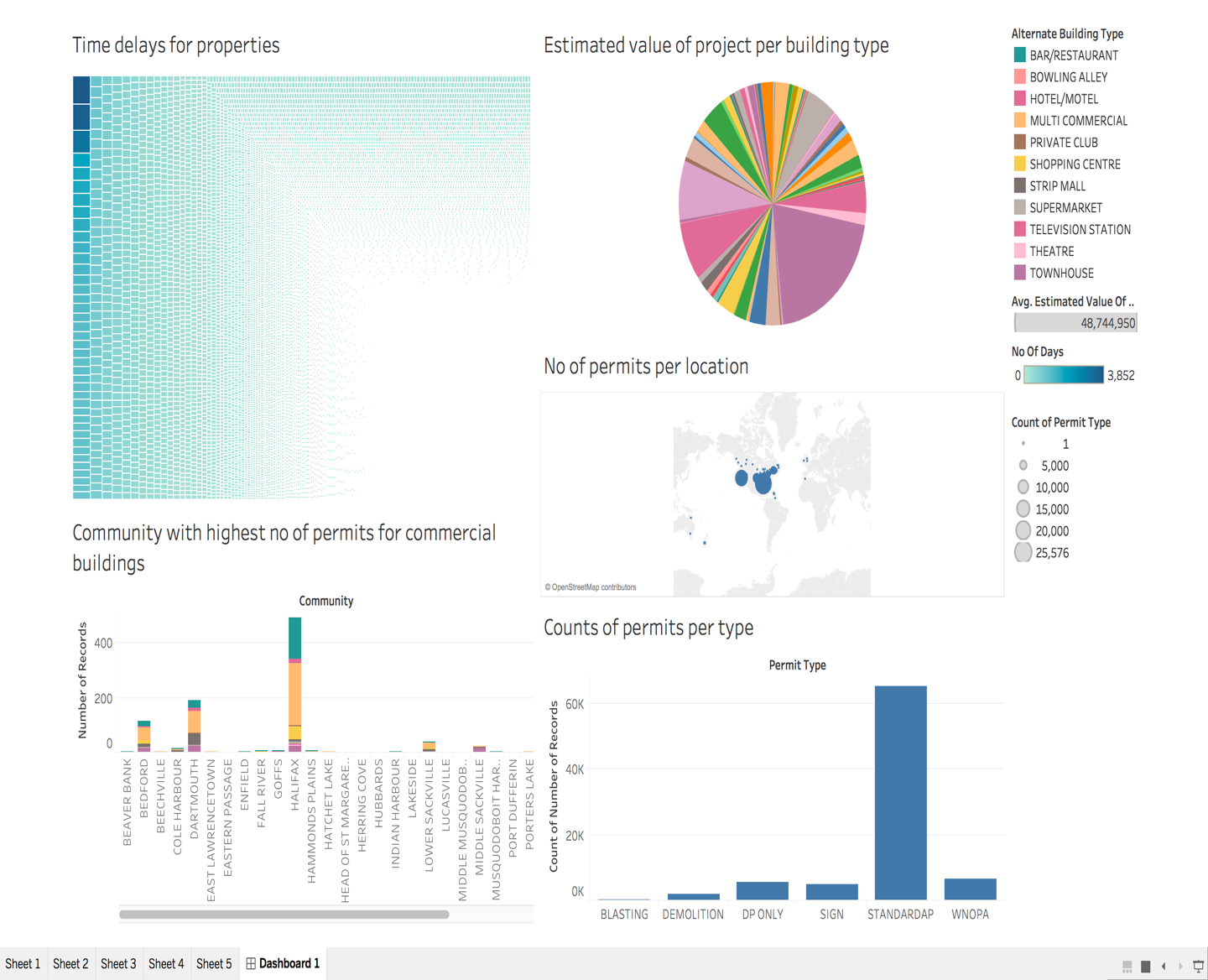


Figure Dashboard with all 5 graphs

1. **Geo Spatial Visualization to plot the number of permits per location.**

In order to plot this graph, we applied group by on X and Y i.e. longitude and latitude. Then we calculated the no of permits for each set of latitude and longitude and size of the circles shown in graph depend on the count of permits for each location.

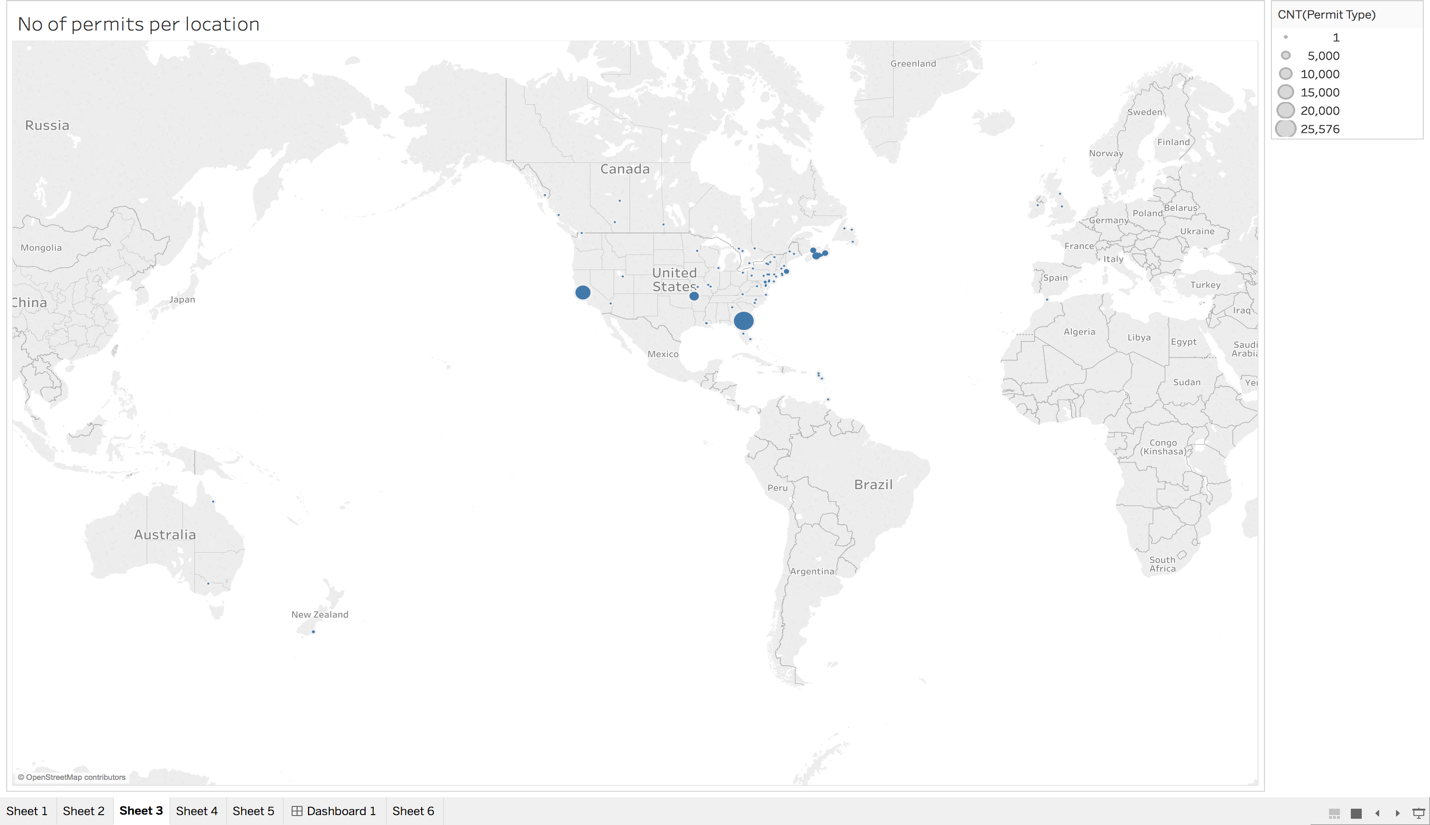


Figure Number of permits per location

1. **Histogram to represent the count of permits per type**

In this graph we try to visualize which permit type has highest number of permit number. So, we first group by permit type and then try to find the count of permit number in each permit type.

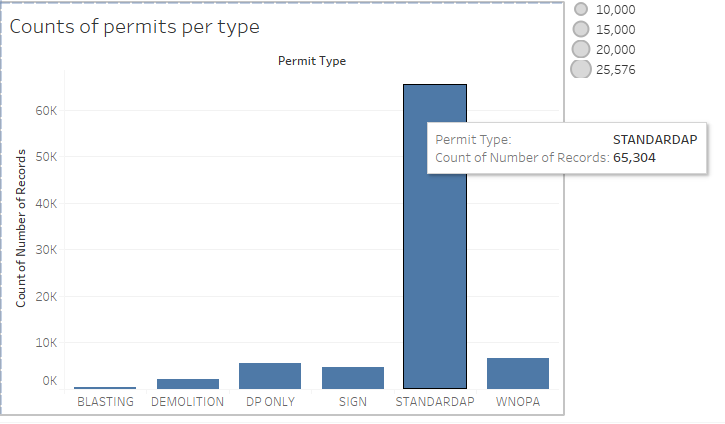
****

Figure Count of permits per type

1. **Graph to identify time gaps between permit applied and permit issuance date and plotting time delays for all permit numbers.**

The properties which tool longest time in getting permit are represented in the graph on the top left corner. In order to plot this graph, we first calculated the difference between two dates i.e. DATE\_OF\_APPLICATION and DATE\_OF\_PERMIT\_ISSUANCE using DATEDIFF in tableau. Then we have plotted the tree map as shown below.

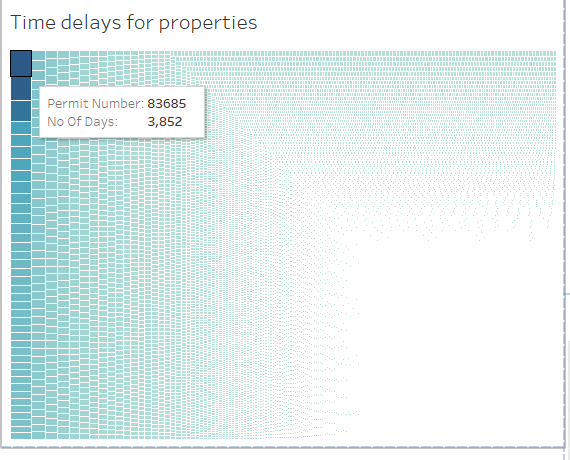


Figure Time gaps between permit applied and permit issuance date

1. **Estimated value of a project and building type where color coding is based on building type.**

In this query we have applied group by on building type and then calculated the average of estimated project cost for each building type.

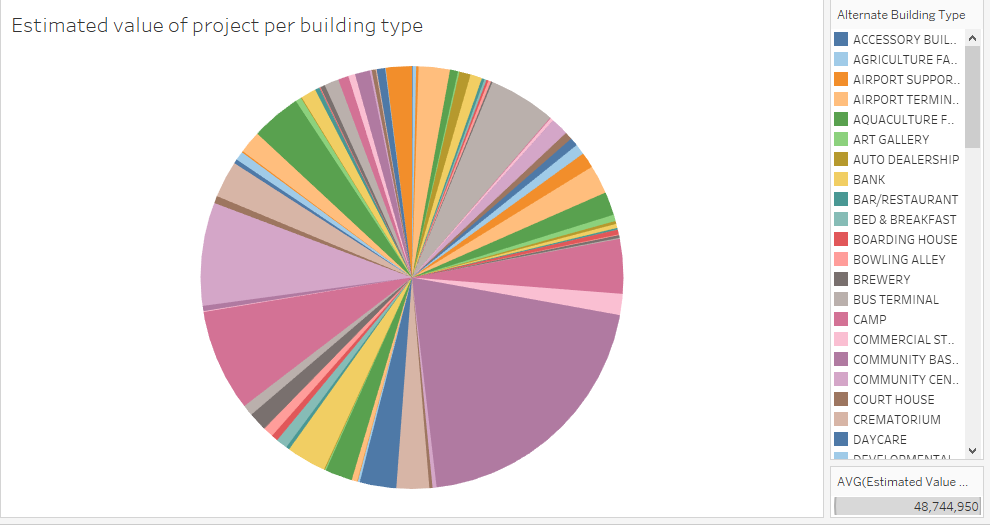


Figure Estimated value of project per building type

1. **In query (4) adding dimension of Community and identifying which community has the highest number of permits issued for commercial buildings.**

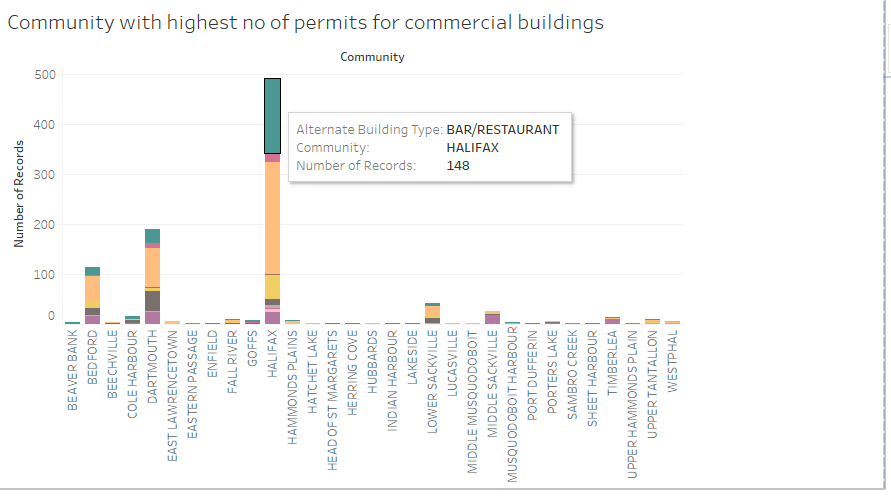


Figure Community with highest number of permits for commercial buildings

1. **OUTPUT AND ANALYSIS**

a) **Geo-Spatial**: The first graph shows the number of permits per location in a geospatial visualization. It has the benefit of showing the maps and number of permit types. It shows analyzed and managed data as per different geographic locations. The higher count of permits in a location can be estimated by the size of the point. The higher the point size shows huge number of permits and smaller point size represents low count of permits. This visualization is easy to gasp for every user and shows all the data with latitude and longitude which makes it easier to analyse and calculate further.

* From the graph we can clearly visualize that united states have higher number of permits than Canada. We see more larger circles and more number of circles in United States than Canada.
* In Canada, we can see that areas in Nova Scotia has more number of permit type than any other area.

b) **Histogram**: This graph is the simplest to analyze as the x-axis represents the Types of Permits and Y-axis represents the Count for the number of records. It helps in visualizing distribution of numerical data over a continuous interval. This visual output shows large amount of data in a clean and smooth diagram. It uses vertical bars to represent data and the height of the bar corresponds to the count of records. The lower the height, lesser the number of records.

* From the graph we can clearly see that STANDARDAP has highest no of permits than other permit types.
* Blasting has the lowest permit number than other Permit types.

c) **Tree-Map**: A treemap chart is used to represent time gaps between permit applied and permit issuance date as it provides a hierarchical view of the data and makes it easy to spot patterns, such as which permit numbers has the longest delays or shortest. The visual shows the tree branches which are in the shape of a rectangle. It is easy to study from the visualization that which permit has the longest time gap and as they are in the top left corners and the bottom right corners of the treemap shows the permit numbers with the shortest time gaps.

We can see that the size of rectangles depends on the span of delay. If delay is more the size is more and it can be clearly visualized. Here are our findings from the graph that we plotted:

* Permit Number 83685 which took 3852 number of days.
* Permit Number 75481 which took 3677 number of days.
* Permit Number 91124 which took 3123 number of days.

d) **Pie Chart**: The pie chart displays the relation between values as well as the relation of a single value to the total. The colour coding makes it easy to compare building types to the total. The arc length of each slice in the visual is proportional to the quantity it represents. The bigger arc depicts the higher estimated value of the project and also compares it to the other projects by looking at the length of the arc.

* From the graph we can see that Library has highest value of average estimated cost than any other building type. It means that average cost in building library is more than any other building.
* In second rank we have Post-Secondary building which has higher average estimated cost than other building type.

e) **Stacked Bars**: This data plotting mechanism shows the highest number of permits issued for commercial buildings with respect to Community. The visual uses bars to show comparisons between different communities and with the ability to breakdown permit count for commercial buildings. Each bar in the visual as a whole represents total permits for each community and segments in the bar depicts the different building type and their number of permits issued.

* From the graph it can be clearly seen that Halifax community has highest no of permits than any other community.
* Strip mall in Dartmouth community is more than the Halifax community. Multi commercial, Bar/Restaurant and Shopping center are more in Halifax than any other community.
* From the visualization we came to conclusion that Halifax has almost every building type and the count is also more in comparison to other communities.

1. **CODE SUBMISSION**

Below is the link for the GitHub repository and all the outputs and files are submitted here.

https://github.com/NK273610/DataWarehouseAssignment4

1. **REFERENCES**
2. **SHEPHERD, A.**

The best business intelligence tool: Tableau vs Microsoft Power BI

**In-text:**(Shepherd, 2018)

**Your Bibliography:**Shepherd, A. (2018). The best business intelligence tool: Tableau vs Microsoft Power BI. Retrieved from http://www.itpro.co.uk/business-intelligence/29132/the-best-business-intelligence-tool-tableau-vs-microsoft-power-bi-1