



Faculty of Computer Science

CSCI 5408 Data Management, Warehousing Analytics

Assignment - 4

**Analyzing Data Patterns Through Effective Visualization
Techniques**

Submitted To

Suhaib Qaiser

Submitted By

Aman Tewary (B00782684)

Nikhil Kamath (B00779777)

Submitted On

July 3, 2018

Table of Contents

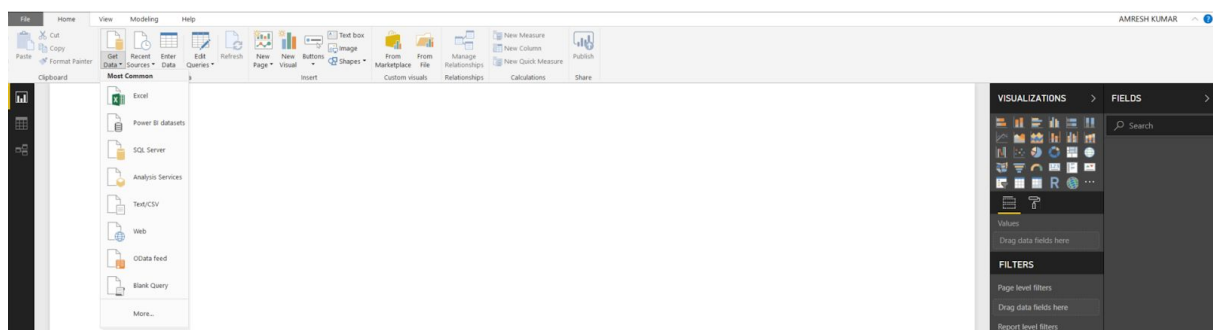
Tool Selection	2
Data Loading	2
Data Cleaning	4
Dashboard	6
Output	7
Code Submission	12
REFERENCE	12

Tool Selection

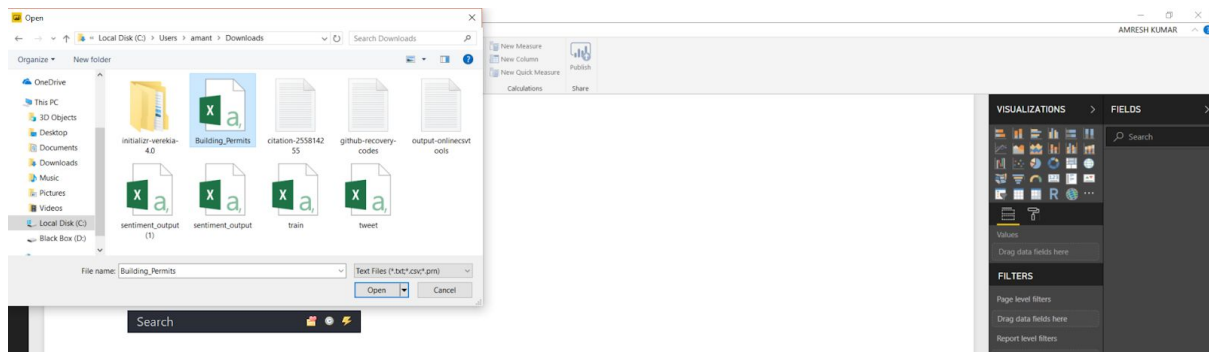
In this assignment we were required to visualize data provided by Halifax Regional Municipality. To complete the required task effectively, we used Microsoft Power BI which is a business analytics service. We chose Power BI because it provides interactive visualizations with self-service business intelligence capabilities. Power BI supports Data Analysis Expression (DAX) which is a library of functions and operators that can be combined to build formulas [1]. We utilized DAX for data cleaning and data manipulation. Power BI also supports real-time filtering of data which we utilized to provide better insight on the given dataset.

Data Loading

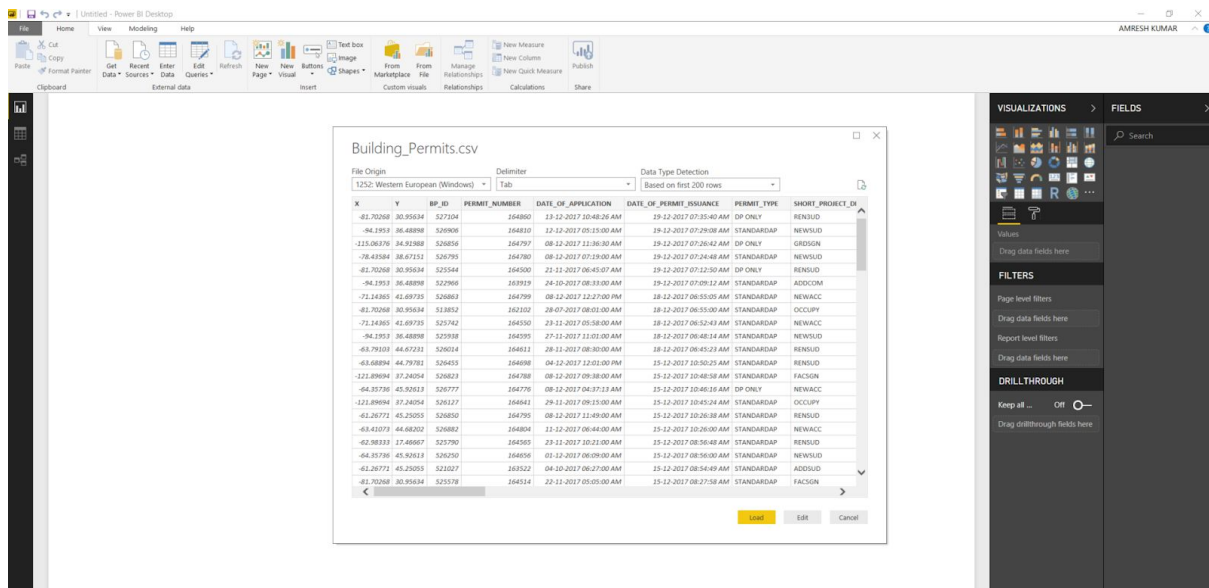
Step 1: From the top panel click on Enter Data and select Text/CSV



Step 2: From Windows Explorer pop-up select the Building_Permits.csv [2] file.



Step 3: The loaded data is automatically formatted into Tab Delimiter CSV file.



Alternatively, user can load data using following query in query editor

```
Csv.Document(File.Contents("C:\Users\amant\Downloads\Building_Permits.csv"),[Delimiter="    ", Columns=30, Encoding=1252, QuoteStyle=QuoteStyle.None])
```

File

Home

Insert

References

Formulas

Developer

View

Help

Clipboard

Font

Paragraph

Styles

Table

Cells

Comments

Navigation

Get Data

Recent Sources

External Data

Refresh

Queries

New

Blank

From Existing Sources

From Data Model

From File

From SharePoint

From Database

From Web

From Other Sources

Text Box

Image

Diagram

Form

Table

Chart

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram

Diagram


Data Cleaning

Step 1: Data is formatted and given its data type using following query.

```
Table.TransformColumnTypes("#Promoted Headers",{{"X", type
number}, {"Y", type number}, {"BP_ID", Int64.Type},
{"PERMIT_NUMBER", Int64.Type}, {"DATE_OF_APPLICATION", type
datetime}, {"DATE_OF_PERMIT_ISSUANCE", type datetime},
{"PERMIT_TYPE", type text}, {"SHORT_PROJECT_DESCRIPTION", type
text}, {"LONG_PROJECT_DESCRIPTION", type text}, {"CIVIC_ID",
Int64.Type}, {"CIVIC_NUMBER", Int64.Type}, {"STREET_NAME", type
text}, {"STREET_TYPE", type text}, {"COMMUNITY", type text},
{"PID", Int64.Type}, {"ELECTORAL_DISTRICT", Int64.Type},
{"MOST_RECENT_INSPECTION", type text}, {"PERMIT_STATUS", type
text}, {"ESTIMATED_VALUE_OF_PROJECT", Int64.Type},
{"ALTERNATE_BUILDING_TYPE", type text}, {"TOTAL_SQ_FOOTAGE",
Int64.Type}, {"SQFT10", Int64.Type}, {"SQFT25", Int64.Type},
{"SQFM30", Int64.Type}, {"SQF030", Int64.Type},
{"EXIST_RES_UNITS", Int64.Type}, {"NEW_RES_UNITS", Int64.Type},
{"NUMBER_OF_STORIES", Int64.Type}, {"INTERNAL_AREA", type text},
{"FID", Int64.Type}})
```

Step 2: After loading and formatting data, Power BI gave an error in following rows because the data type of Permit_Number column is whole number, however, these three rows contained string values.

1.2 Row Number	1.2 X	1.2 Y	1.3 BP_ID	1.3 PERMIT_NUMBER	DATE_OF_APPLICATION	DATE_OF_PERMIT_ISSUAN...	1.4 PERMIT_TYPE	1.4 SHORT_PROJECT_DES
1	71007	-121.89694	37.24054	55303 Error	null	15-02-2005 07:23:36 AM	WNOPA	REN2UD
2	71184	-121.89694	37.24054	55024 Error	null	09-02-2005 05:30:14 AM	WNOPA	RENSUD
3	71355	-121.89694	37.24054	54840 Error	null	04-02-2005 07:41:00 AM	WNOPA	RENSUD



To clean the data we replaced the “Error” with random Permit_Number using following query. (It should noted that to avoid any issue with visualization, these three rows were filtered out while plotting.)

```
Table.ReplaceValue("#Sorted
Rows", "Error", "64468", Replacer.ReplaceValue, {"PERMIT_NUMBER"})
```

Step 3: The ALTERNATE_BUILDING_TYPE column contained 71942 blank fields.

To rectify this we replaced blank fields with “Unknown” as no concrete data was available to verify the content of the field. (It should noted that to avoid any issue with visualization, “Unknown” building type was filtered out.)

```
Table.ReplaceValue("#Sorted
Rows", "", "Unknown", Replacer.ReplaceValue, {"ALTERNATE_BUILDING_TYPE
"})
```

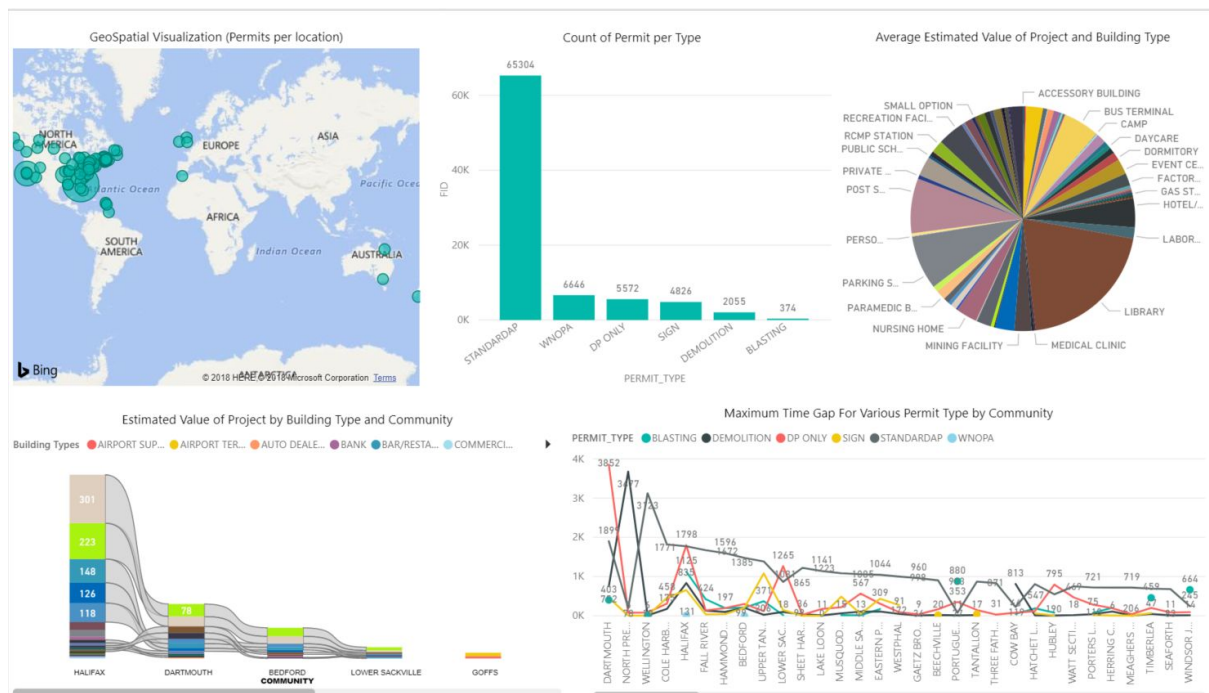
Step 4: As we were required to plot the difference in permit application date and permit issuance date, we created a column with difference in days using following query.

```
Diff_Permit_Application_Date_(DAY) =
DATEDIFF(Building_Permits[DATE_OF_APPLICATION], Building_Permits[DA
TE_OF_PERMIT_ISSUANCE], DAY)
```

Dashboard

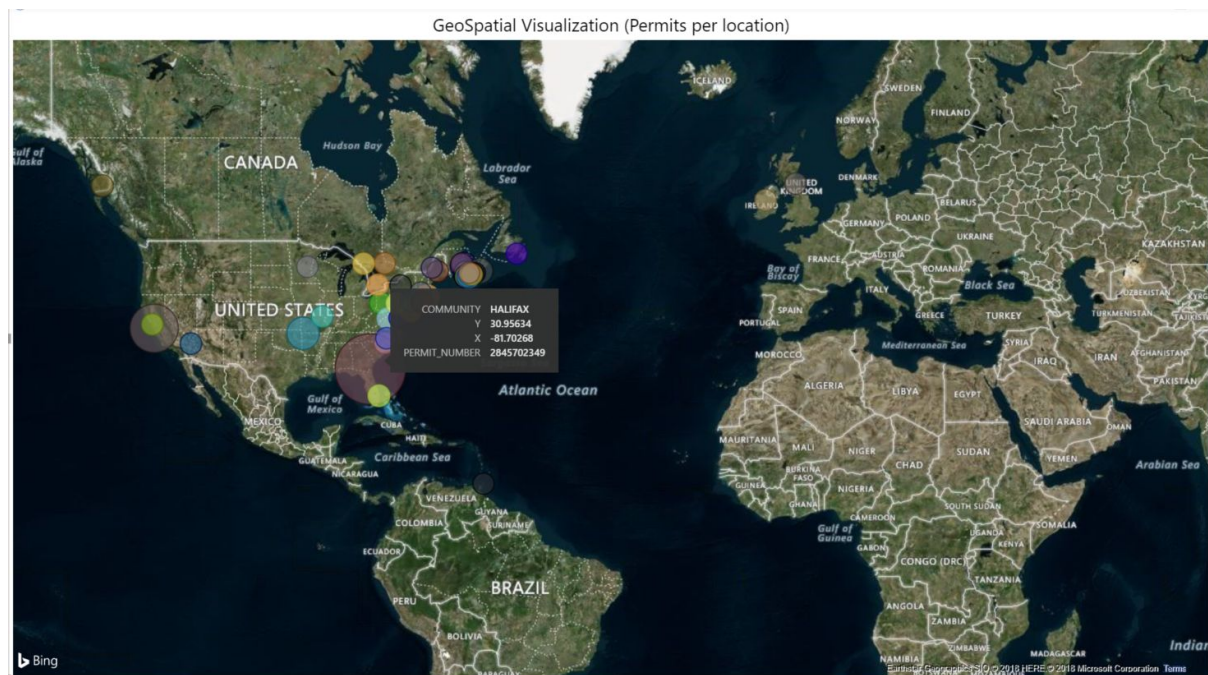
URL:

<https://app.powerbi.com/view?r=eyJrIjoiaZGQzNGFhZjktOTQxNS00OTdhLTlkZDMtMTBhYTRhYzBIMmU4IiwidCI6ImQ3OTA5NTVjLTc5MDMtNDc1NC04NDJiLTMyNTAzZDliNmVkYiIsImMiOjEwfQ%3D%3D>



Output

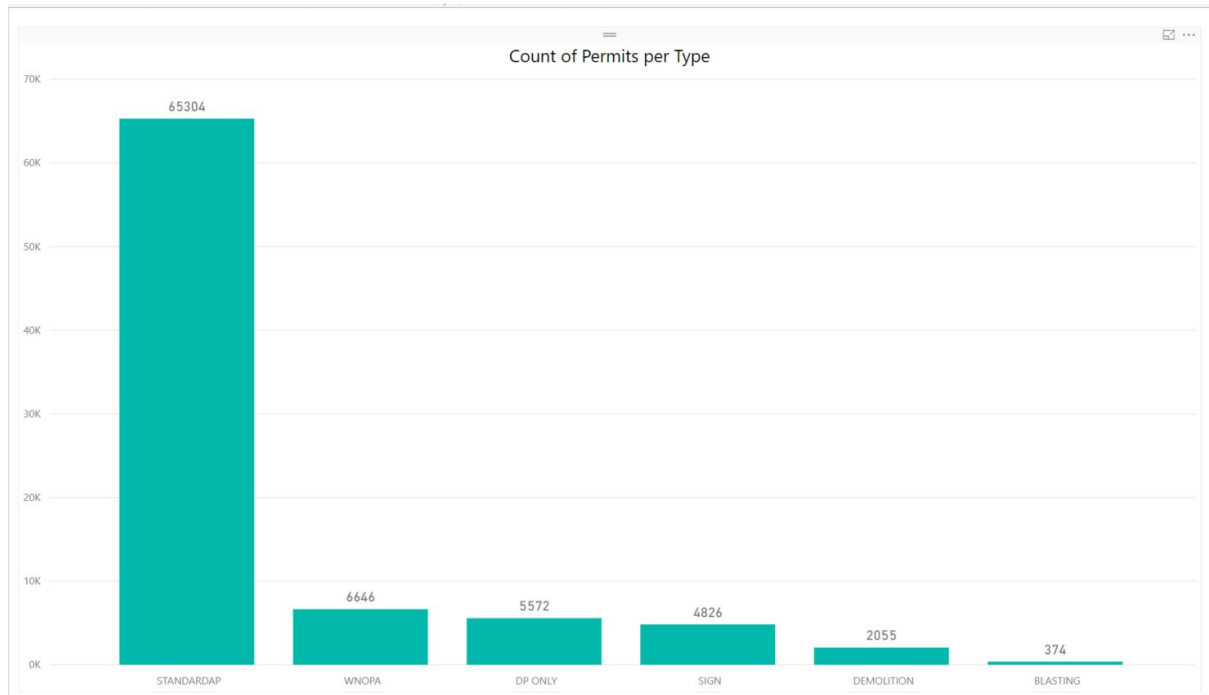
1. GeoSpatial Visualization



Analysis:

Total number of permit is 84,777 across all 192 communities. The distribution ranges from one (Beaver Dam, Chaplin and Ecum Secum) to 25,576 (Halifax), a difference of 25,575, averaging 441.55. The distribution is positively skewed as the average of 441.55 is much greater than the median 45. Total number of permits is highly concentrated with just 19 of the 192 communities (9.9%) representing 83% of the total. Halifax accounts for 30% of overall number of permits. Halifax (25,576) is almost 58 times bigger than the average across the 192 communities. Dartmouth and Bedford stood out with very high number of permits.

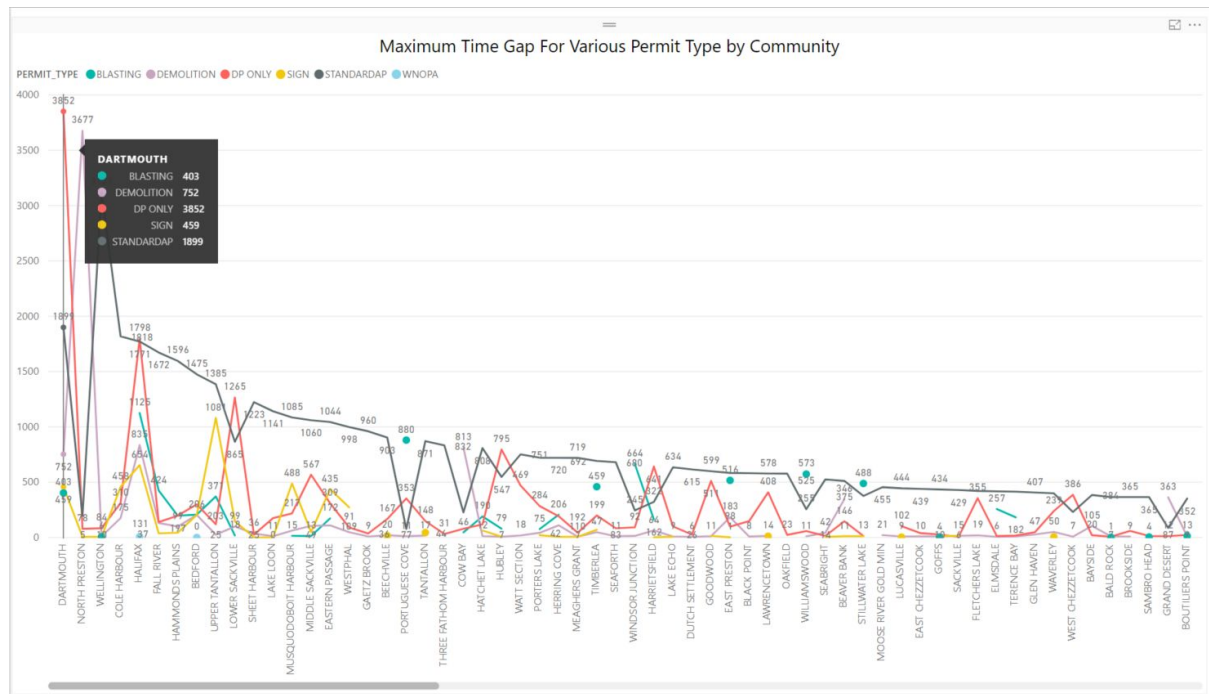
2. Count of Permit per Type



Analysis:

There are total 84,777 permits across all six permit types. The minimum value is 374 (Blasting) and the maximum is 65,304 (Standardap), a difference of 64,930, averaging 14,130. The count of permits is highly concentrated with just one of the six types (17%) representing 77% of the total. Standardap accounts for 77% of total number of permits. Standardap (65,304) is almost five times bigger than the average across the six types.

3. Maximum Time Gap For Various Permit Type



Analysis:

We plotted the maximum days difference in permit application date and permit issuance date by Community and by Permit Type. We filtered out communities with less than 5 days difference. The sum of maximum of difference in permit application date and permit is 101608 days across all 177 communities. The distribution ranges from 12 days (Brookvale) to 7365 days(Dartmouth), a difference of 7 7353 days, averaging 574 days. The top six communities represent over a quarter (26%) of overall value. The top 3 communities based on total maximum days difference across all permit types are discussed in details below:

Dartmouth:

The sum of maximum days difference is 7365 across all five permits (representing 7.25% of total sum of maximum difference across all 177 communities.) The minimum value is 403(Blasting) and the maximum is 3852 (DP Only) which is almost three times bigger than the average across the five permits.

Halifax:

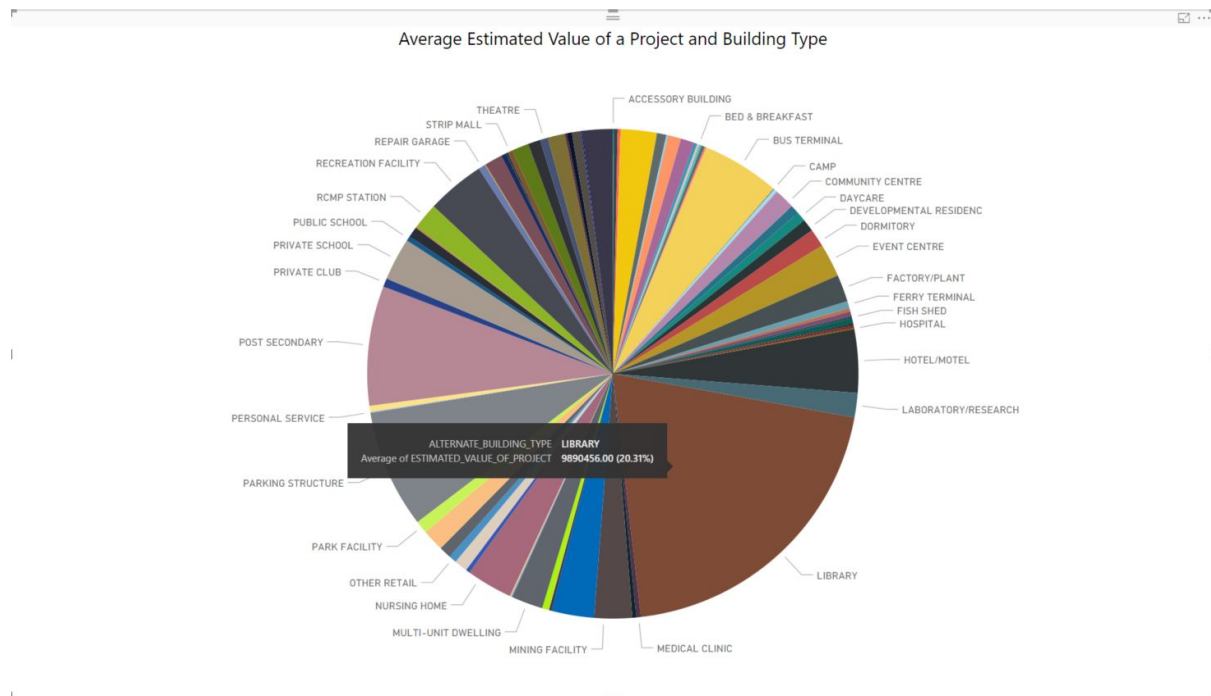
The sum of maximum days difference is 6191 across all six permits (representing 6.09% of total sum of maximum difference across all 177 communities.) The minimum value is 8(WNOPA) and the maximum is 1798(DP Only).

North Preston:

The sum of maximum days difference is 3956 across all four permits (representing

3.89% of total sum of maximum difference across all 177 communities.) The minimum value is five(SIGN) and the maximum is 3677 (Demolition).

4. Average Estimated Value of a Project by Building Type



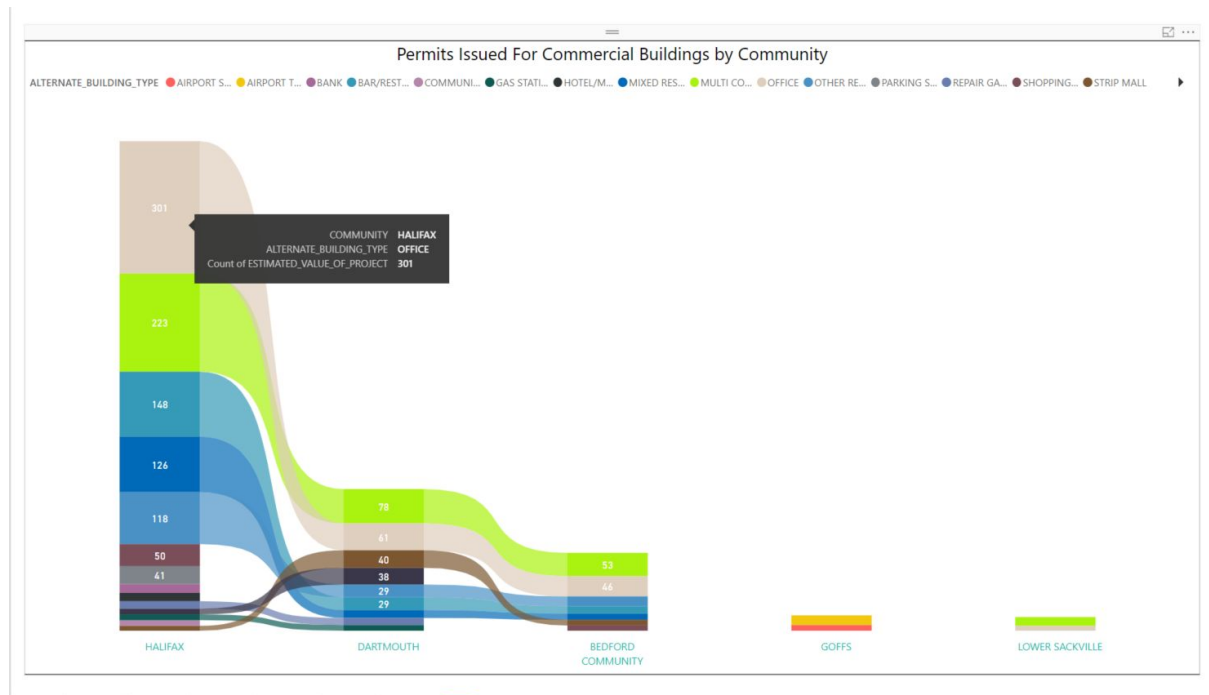
Analysis:

In the given dataset, there are 83 types of building.

However, we have dropped “Unknown” type as it was added during data cleaning by replacing blank fields.

The sum of average of estimated value of project is \$48.7 million across all 81 types of buildings. The distribution ranges from \$3,750 (Community Based Option) to \$9.9 million (Library), a difference of \$9.9 million, averaging \$601,291. The distribution is positively skewed as the average of \$601,291 is much greater than the median of \$205,191. Average of Estimated value of project is very concentrated with 77% of total represented by just 19 of 81 types of buildings (23%). Post Secondary and Parking Structure were outliers with very high Average of Estimated value of project.

5. Permit Issued For Commercial Buildings by Community.



Analysis:

We added Community with the building type and count of estimated value of project. We filtered out all the non-commercial buildings. We added a filter in the count of Estimated value of project to retrieve only values greater than 10 as there was significant gap in distribution of data.

The total number of permits for commercial buildings is 1791 across all five communities. The distribution ranges from 31 (Lower Sackville) to 1197 (Halifax), a difference of 1166, averaging 358.2. Halifax (1197) is more than three times bigger than the average across the five communities. Halifax has the highest number of permits issued for commercial buildings. It is discussed below in more details:

Halifax:

Total number of permits for commercial building is 1197 across all 15 building types (representing 67% of total number of permits). The distribution ranges from 11 (Strip Mall) to 301 (Office), a difference of 290, averaging 79.8. The distribution is positively skewed as the average of 79.8 is much greater than the median 41. Halifax is relatively concentrated with 84% of the total represented by six of the 15 building types. Office accounts for a quarter of overall number of permits (301) which is almost four times bigger than the average across the 15 commercial building types.

Code Submission

GITHUB URL:

<https://github.com/amantewary/Analyzing-Data-Patterns-Through-Effective-Visualization-Techniques-Power-BI.git>

Dashboard URL:

<https://app.powerbi.com/view?r=eyJrljoiZGQzNGFhZjktOTQxNS00OTdhLTlkZDMtMTBhYT RhYzBI MmU4IiwidCI6ImQ3OTA5NTVjLTc5MDMtNDc1NC04NDJiLTMyNTAzZDliNmVkYiIsImMiOiJEWfQ%3D%3D>

REFERENCE

- [1] "Introduction to DAX - Power BI." [Online]. Available: <https://docs.microsoft.com/en-us/power-bi/guided-learning/introductiontodax?tutorial-step=1>. [Accessed: 02-Jul-2018]
- [2] ".Building Permits" [Online]. Available: <https://catalogue-hrm.opendata.arcgis.com/datasets/building-permits>. [Accessed: 02-Jul-2018]