Table of Contents

[Task 1 – Identify the Data Set 3](#_Toc27142650)

[Task 2 – Plan the Implementation of BI System 4](#_Toc27142651)

[2.1 Business Model 4](#_Toc27142652)

[2.2 Business rule 4](#_Toc27142653)

[2.2.1 Host 4](#_Toc27142654)

[2.2.2 Guest 5](#_Toc27142655)

[2.2.3 New York Airbnb 5](#_Toc27142656)

[2.3 Information flow 6](#_Toc27142657)

[2.4 Analytics spectrum questions 7](#_Toc27142658)

[2.5 Data architecture 7](#_Toc27142659)

[2.5.1 Data modelling 7](#_Toc27142660)

[2.5.2 Data preparation 8](#_Toc27142661)

[2.5.3 Data integration 9](#_Toc27142662)

[2.5.4 Data Profiling 10](#_Toc27142663)

[2.5.5 Data quality 10](#_Toc27142664)

[2.6 The location intelligence technique 11](#_Toc27142665)

[Task 3 – Carry Out and Evaluate the Technologies that make up BI 12](#_Toc27142666)

[3.1 Analytics spectrum questions 12](#_Toc27142667)

[3.1.1 Top 15 expensive listing of all room types 12](#_Toc27142668)

[3.1.2 Top 15 cheapest listing of all room types 13](#_Toc27142669)

[3.1.3 How many available rooms for 365 days for all places and rooms? 13](#_Toc27142670)

[3.1.4 What is the percentage room types (private, entire, shared) exist in this location? 14](#_Toc27142671)

[3.1.5 How many room\_type (private, entire, shared) existed for each of the neighbourhood group? 15](#_Toc27142672)

[3.1.6 Does the room types of a specific neighbourhood group affect the amount of price? 16](#_Toc27142673)

[3.1.7 Predict the next room type’s price by calculating the average price of it? 18](#_Toc27142674)

[3.1.8 Predict the each of the room type’s price for each neighbourhood\_group. 19](#_Toc27142675)

Table

[Table 1.0 Business rule for host 4](#_Toc27134530)

[Table 2.0 Business rule for guest 5](#_Toc27134531)

[Table 3.0 Business rule for New York Airbnb 5](#_Toc27134532)

Figures

[Figure 1.0 Information flow for New York Airbnb 6](#_Toc27142754)

[Figure 2.0 ERD for New York Airbnb 7](#_Toc27142755)

[Figure 3.0 Filter and Replace Data process 8](#_Toc27142756)

[Figure 4.0 Filters Data process 8](#_Toc27142757)

[Figure 5.0 Replace Missing Value process 9](#_Toc27142758)

[Figure 6.0 Generate Attributes process 9](#_Toc27142759)

[Figure 7.0 Generate Attributes process generate new attribute ”price\_range\_from\_1\_to\_10000” 9](#_Toc27142760)

[Figure 8.0 Algorithms for new attribute ”price\_range\_from\_1\_to\_10000” 10](#_Toc27142761)

[Figure 9.0 Mapping for price range 1-10000 11](#_Toc27142762)

[Figure 10.0 Line graph of Top 15 expensive listing of all room types 12](#_Toc27142763)

[Figure 11.0 Line graph of Top 15 cheapest listing of all room types 13](#_Toc27142764)

[Figure 12.0 Table of Total available rooms for 365 days 13](#_Toc27142765)

[Figure 13.0 Piechart for room types 14](#_Toc27142766)

[Figure 14.0 Bar chart of room type existed for each neighbourhood group 15](#_Toc27142767)

[Figure 15.0 Point Mapping for room types 15](#_Toc27142768)

[Figure 16.0 Bar chart of the relationship between room types of a specific neighbourhood group and the price 16](#_Toc27142769)

[Figure 17.0 Bar chart of the average price of each room type 18](#_Toc27142770)

[Figure 18.0 Bar Chart of Neighbourhood for each room type’s price 19](#_Toc27142771)

[Figure 19.0 Table of Neighbourhood for each room type’s price from excel file 19](#_Toc27142772)

# Task 1 – Identify the Data Set

New York City Airbnb Open Data by Dgomonov from Kaggle <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>. This data set has a few attribute such as

* Listing\_ID,
* Listing\_Name
* Host\_ID
* Host\_Name
* Neighbourhood\_Group
* Neighbourhood
* Latitude
* Longitude
* Room\_Type
* Price
* Minimum\_Nights
* Number\_of\_Reviews
* Last\_review
* Reviews\_per\_Month
* Calculated\_Host\_Listing\_Count
* Availability\_365.

# Task 2 – Plan the Implementation of BI System

## 2.1 Business Model

Brokerage business model is a one of the Airbnb business models where a third party helps connect the sellers and buyers to facilitate a transaction [1]. The business usually charges a transaction fee or service fee for their services. Most real estate business use the brokerage business model. By using this business model, the organizations have the ability to create a market opportunity and also the ability to create a powerful barrier of their network for an established broker firm but newer organizations that use this business model could face difficulties to establish the network due to the effect of established brokerage firms.

## 2.2 Business rule

### 2.2.1 Host

Table 1.0 Business rule for host

|  |  |
| --- | --- |
| Rules | Description |
| Pay taxes | Host must pay taxes to their cities for this extra income stream |
| Obtain a permit or a license | Host must register a permit or a license to operate a rental |
| Check subletting is permitted | Host must check their lease agreement and see if subleasing is permitted |
| Check city’s laws | Host must check their city’s laws that how many days per year they can sublet |
| Booking modifications | Host must refund the full amount to the guest if the host cancel the confirmed booking |
| Set listing fee | Host must set a fixed price for your listing |

### 2.2.2 Guest

Table 2.0 Business rule for guest

|  |  |
| --- | --- |
| Rules | Description |
| Member verification | Guest must be at least 18 years old in order to access and use the Airbnb Platform |
| Only Registered Guest are allowed | All guest must be registered firsthand |
| Damage accommodations | Guest must leave the accommodations conditions as there are in the first place |
| Illegal activity | Guest are strictly prohibited to do any illegal activity |
| Booking modifications | Guest can cancel a confirmed booking at any time pursuant to the Listing’s cancellation policy set by the host and the Airbnb Payments will refund the amount of the Total Fees |
| Pay deposit | Guest must pay deposit upon booking |

### 2.2.3 New York Airbnb

Table 3.0 Business rule for New York Airbnb

|  |  |
| --- | --- |
| Rules | Description |
| Charge services fee | Airbnb may charge fees to Host or Guest |

## 2.3 Information flow

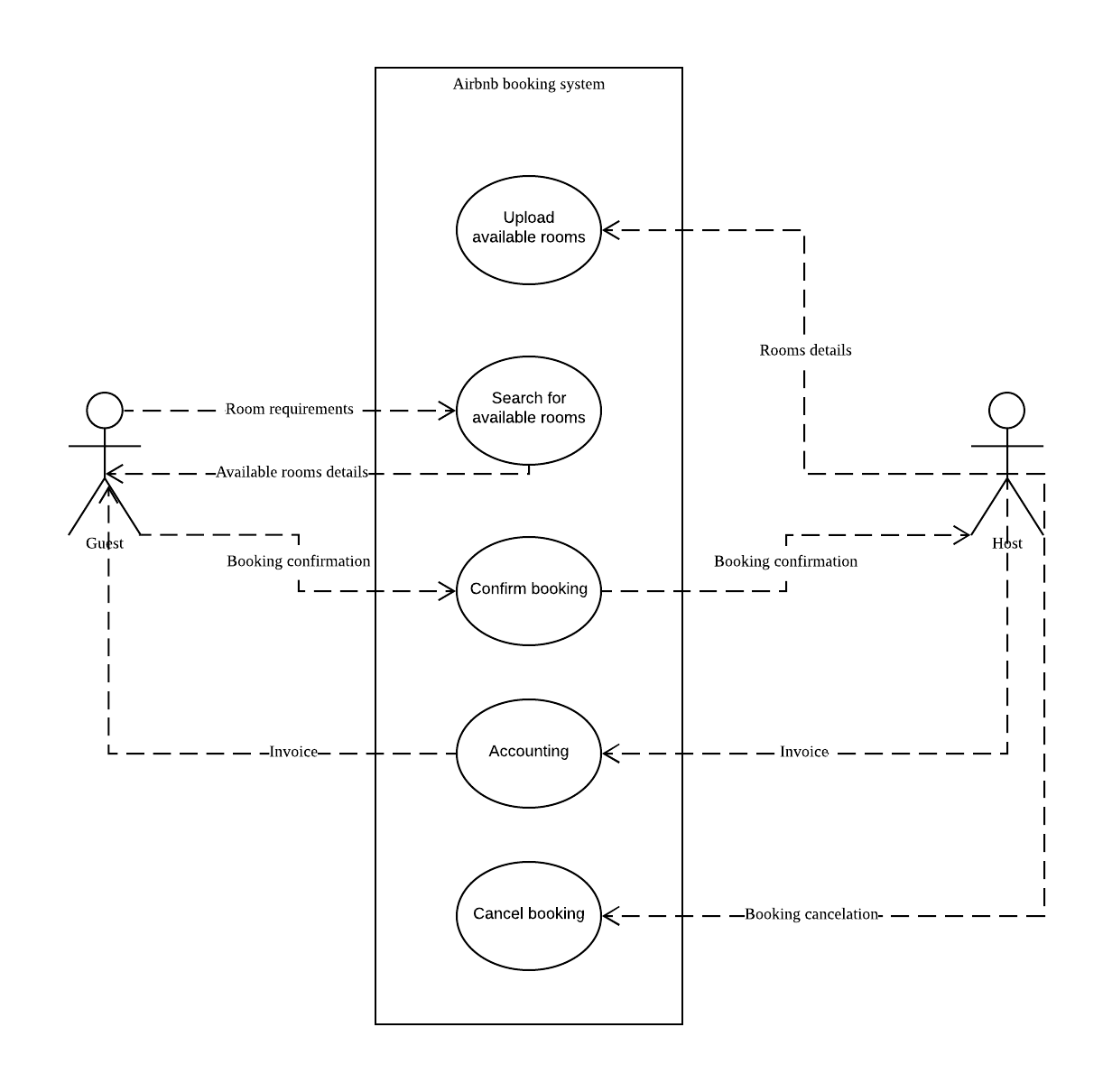


Figure 1.0 Information flow for New York Airbnb

## 2.4 Analytics spectrum questions

* Top 15 expensive listing of all room types.
* Top 15 cheapest listing of all room types.
* How many available rooms for 365 days for all places and rooms?
* What is the percentage room types (private, entire, shared) exist in this location?
* How many room\_type (private, entire, shared) existed for each of the neighbourhood\_group?
* Does the room type of a specific neighbourhoodgroup affect the amount of price?
* Predict the next room type’s price by calculating the average price of it?

## 2.5 Data architecture

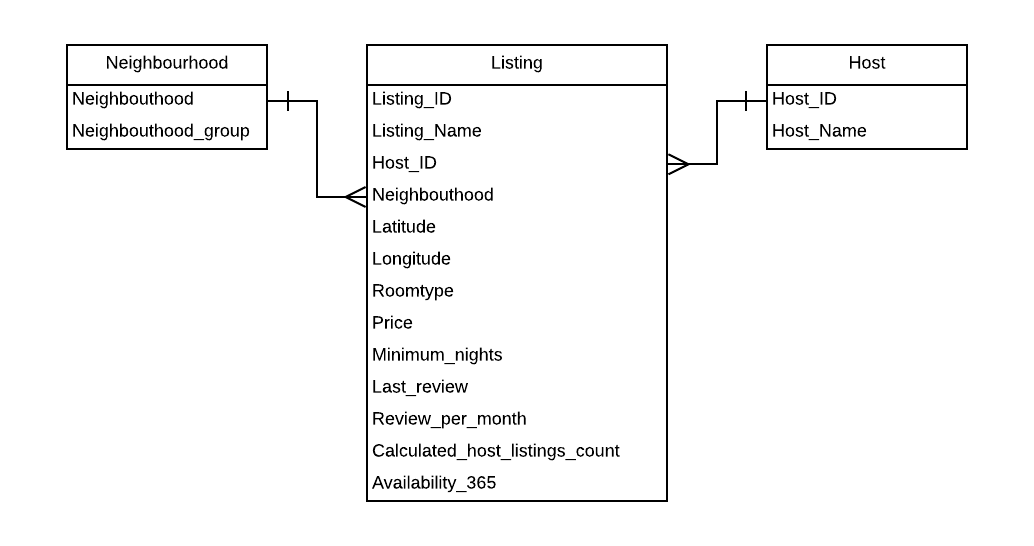
2.5.1 Data modelling 

Figure 2.0 ERD for New York Airbnb

### 2.5.2 Data preparation

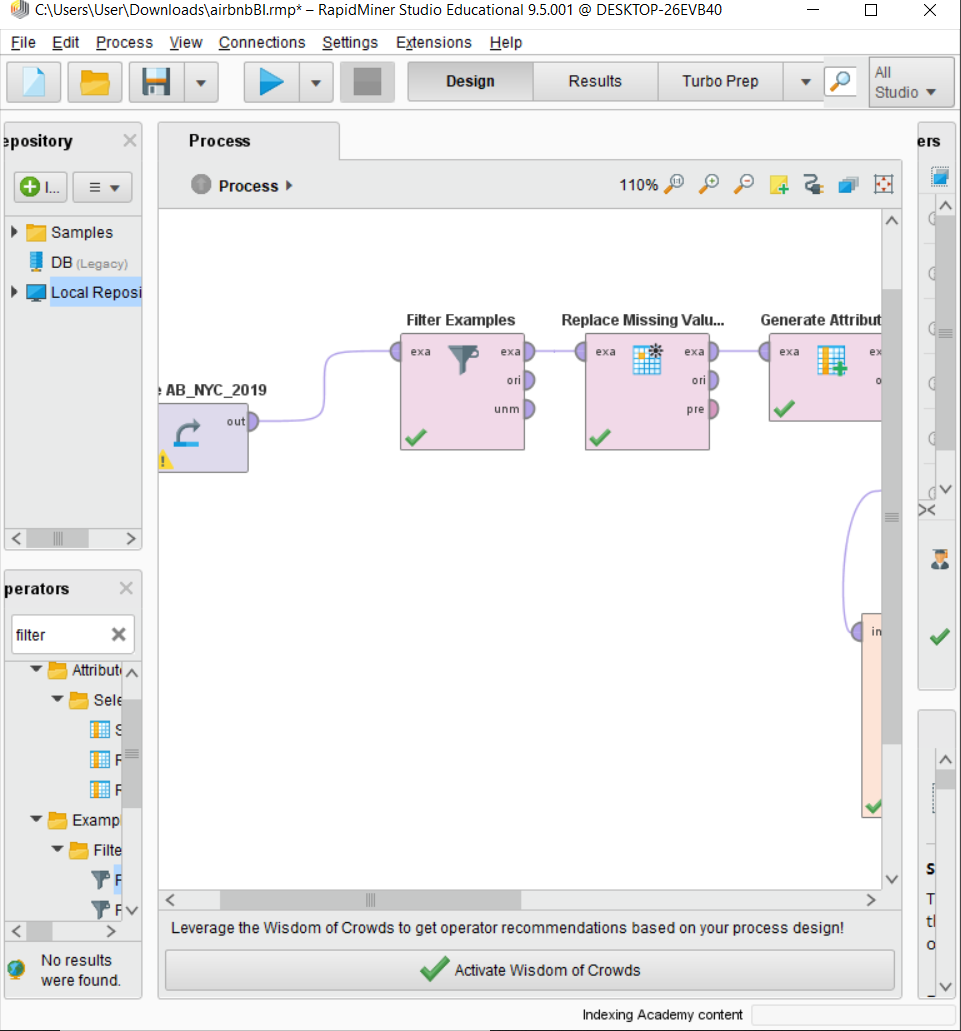


Figure 3.0 Filter and Replace Data process

Filter Examples process and Replace Missing Value process are used in this case to clean and transform the raw data to clean and processed data.

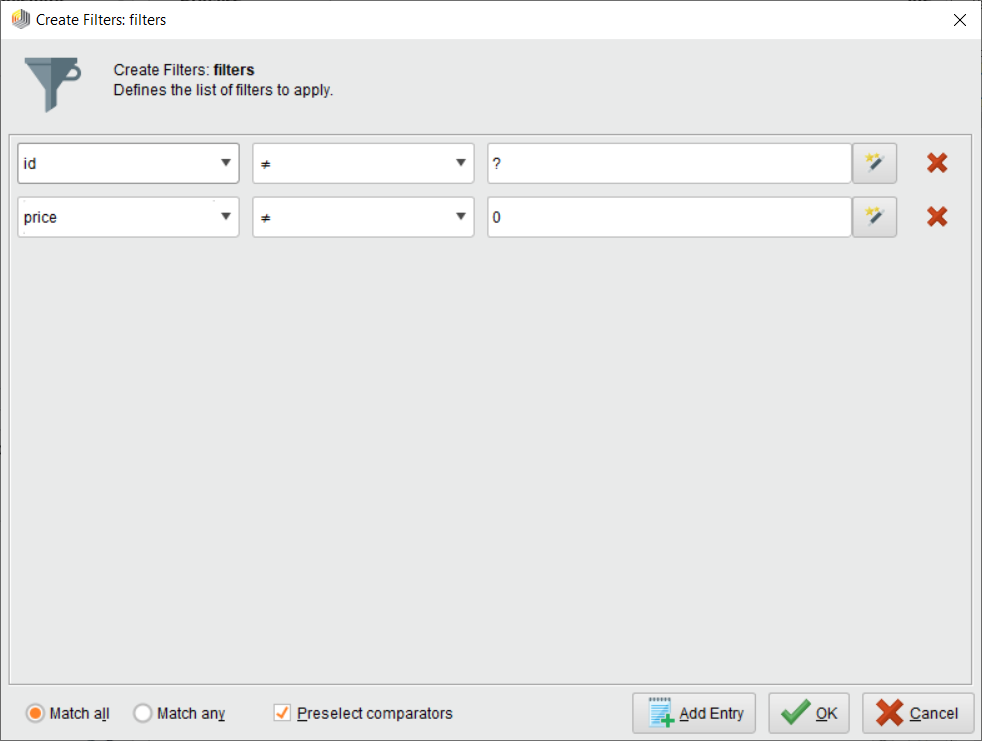


Figure 4.0 Filters Data process

Filter Examples process filter attribute Listing\_ID from error value which is ‘?’ and filter attribute Price from zero value.

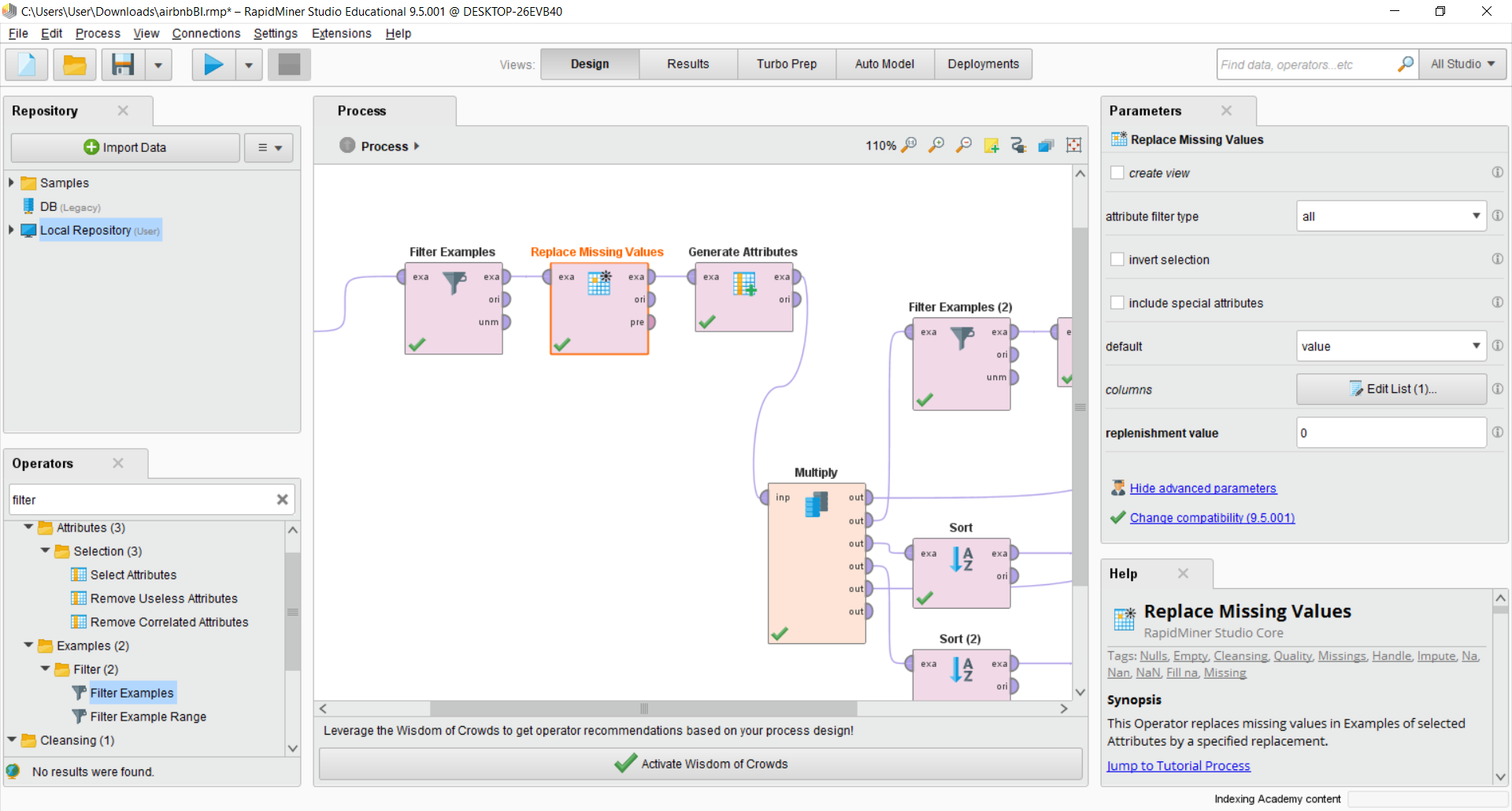


Figure 5.0 Replace Missing Value process

Replace Missing Value process replace value that consist missing value and error value to 0.

### 2.5.3 Data integration

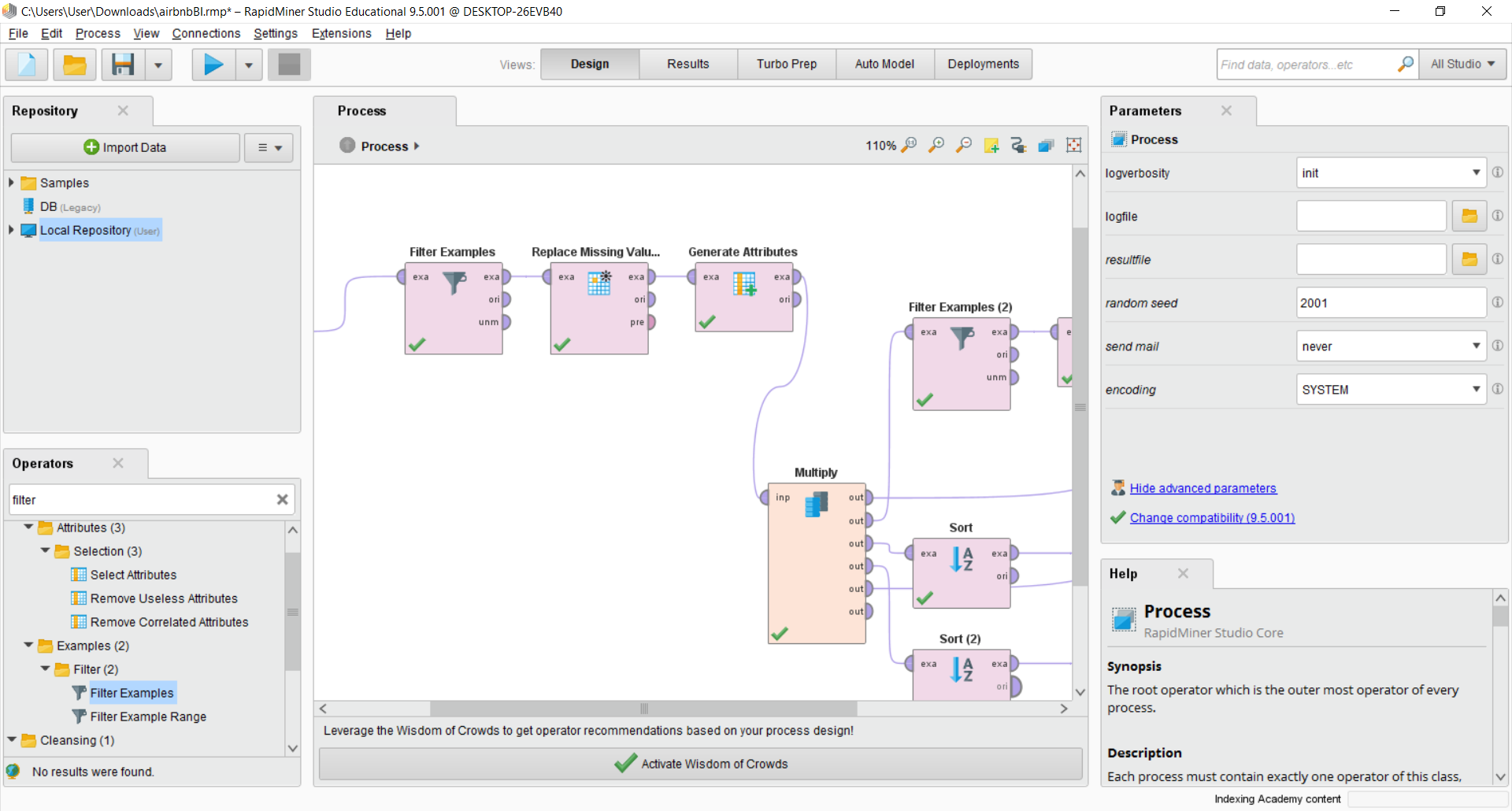


Figure 6.0 Generate Attributes process

Generate Attributes process generate new attribute and integrate with the existing data.

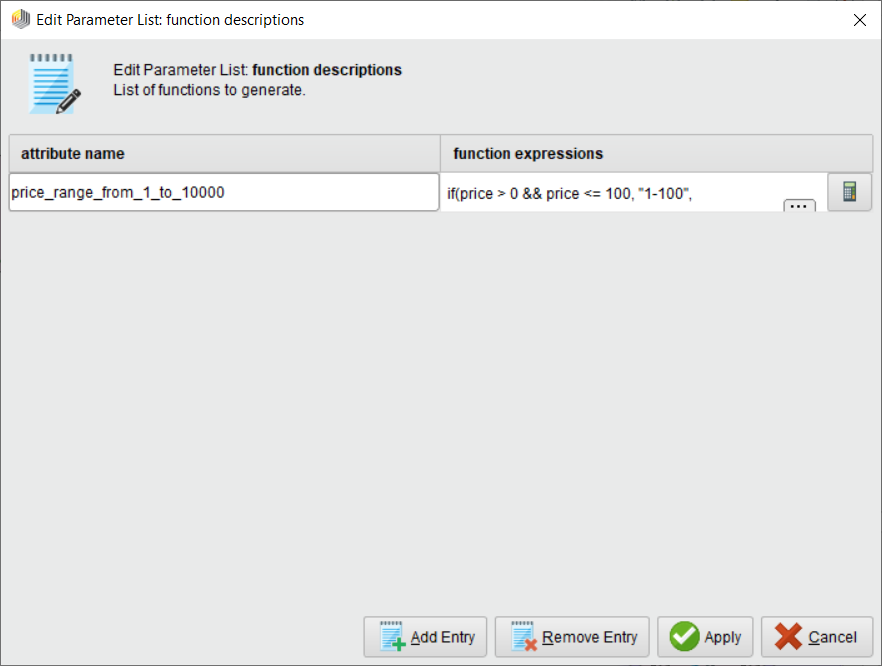


Figure 7.0 Generate Attributes process generate new attribute ”price\_range\_from\_1\_to\_10000”

Generate Attributes process generate a new attribute called “price\_range\_from\_1\_to\_10000”.

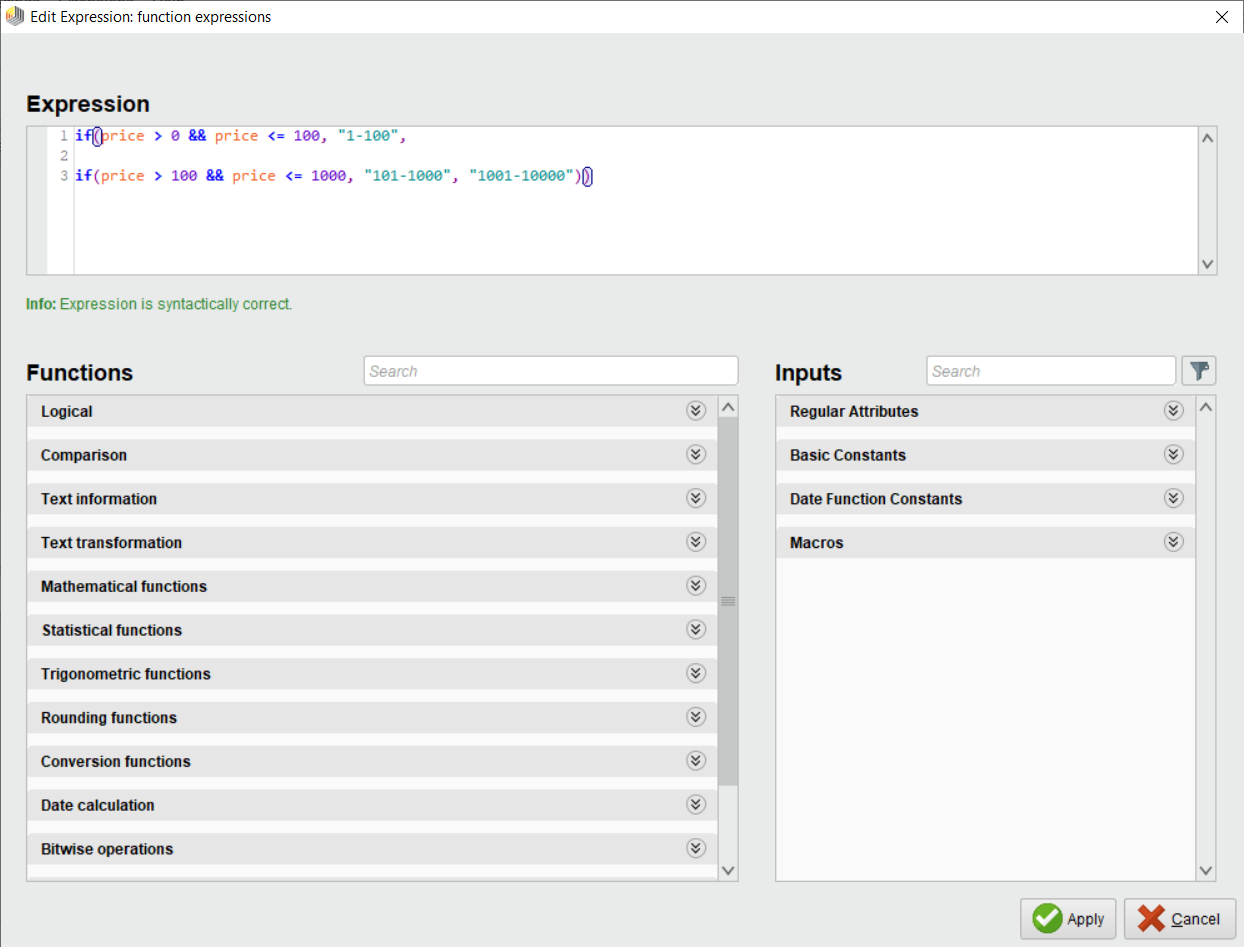


Figure 8.0 Algorithms for new attribute ”price\_range\_from\_1\_to\_10000”

Generate Attributes process use this algorithm to generate the new attribute.

### 2.5.4 Data Profiling

Data is examining from an existing information source and collected statistics for it.

### 2.5.5 Data quality

Data is proved quality as it has been clean and transform at data preparation. Data no missing value, no error value and now is clean and quality.

## 2.6 The location intelligence technique

RapidMiner is used as a software to perform such technique in order to produce the result from the analytics spectrum questions. For example mapping to show the price range from 1 to 10000.

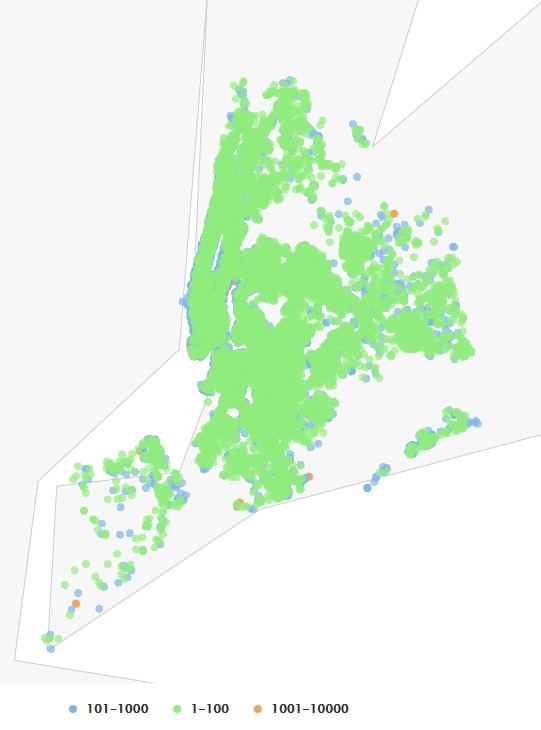


Figure 9.0 Mapping for price range 1-10000

# Task 3 – Carry Out and Evaluate the Technologies that make up BI

## 3.1 Analytics spectrum questions

### 3.1.1 Top 15 expensive listing of all room types

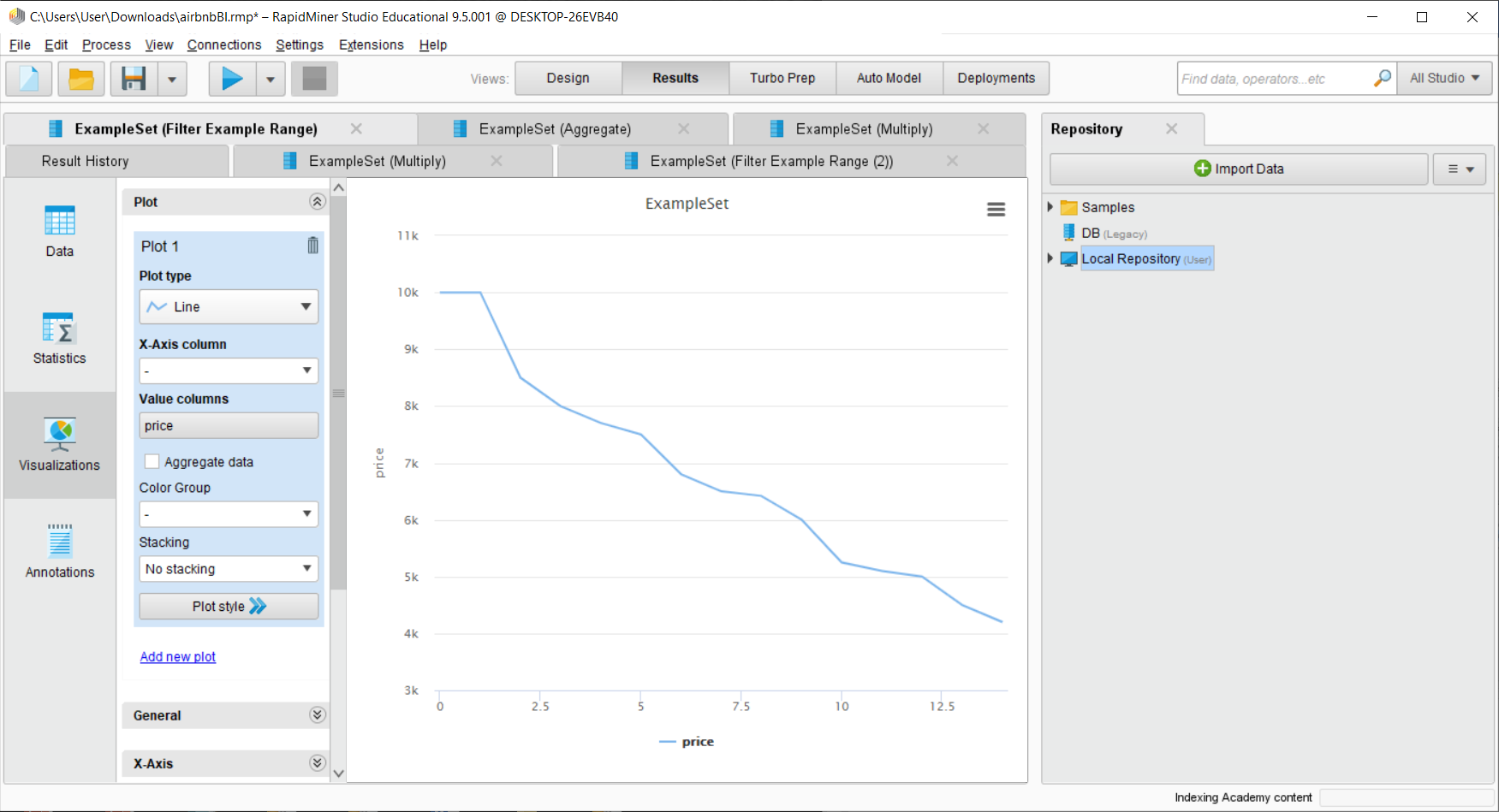


Figure 10.0 Line graph of Top 15 expensive listing of all room types

Conclusion:

Top 15 expensive listing of all room types are plotted on the graph. The 15 most expensive listing are ranging from USD 4200-10000.

### 3.1.2 Top 15 cheapest listing of all room types

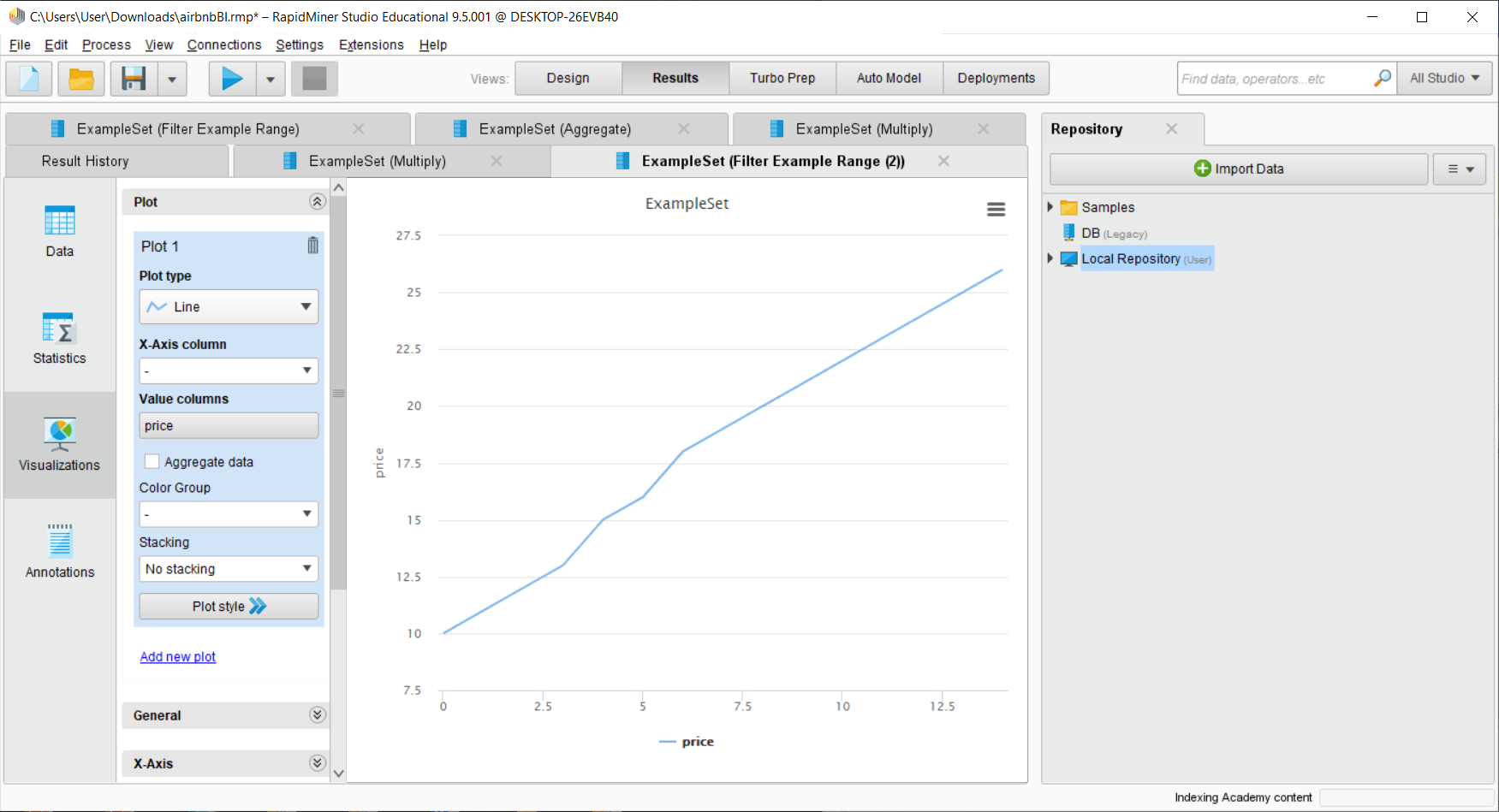


Figure 11.0 Line graph of Top 15 cheapest listing of all room types

Conclusion:

Top 15 cheapest listing of all room types are plotted on the graph. The 15 most cheapest listing are ranging from USD 10-26.

### 3.1.3 How many available rooms for 365 days for all places and rooms?

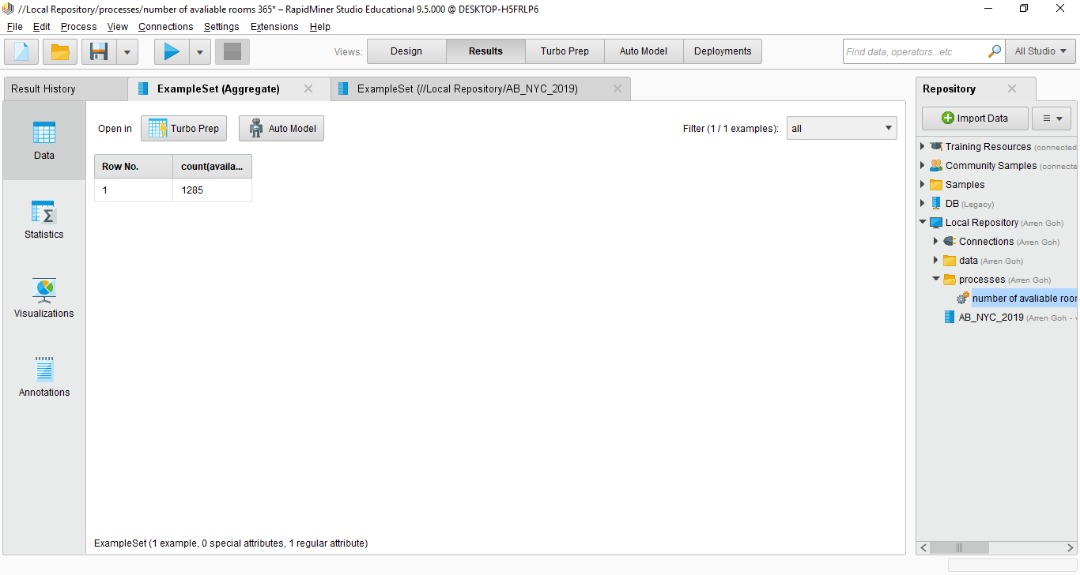


Figure 12.0 Table of Total available rooms for 365 days

Conclusion:

The total available rooms is record on the table. The total available rooms are 1285.

### 3.1.4 What is the percentage room types (private, entire, shared) exist in this location?

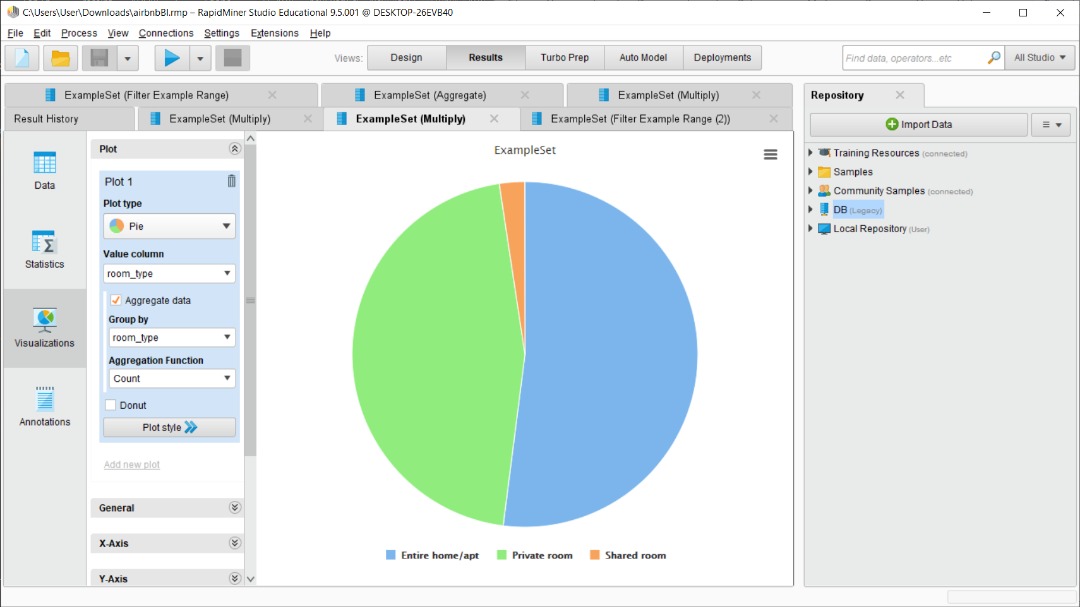


Figure 13.0 Piechart for room types

Conclusion:

All amount of the room types are plotted on the piechart. The total room types are 48664. The entire home/apt room types are 25319 which is 52%. The private room types are 22191 which is 45.6%. The shared room types are 1154 which is 2.4%.

### 3.1.5 How many room\_type (private, entire, shared) existed for each of the neighbourhood group?

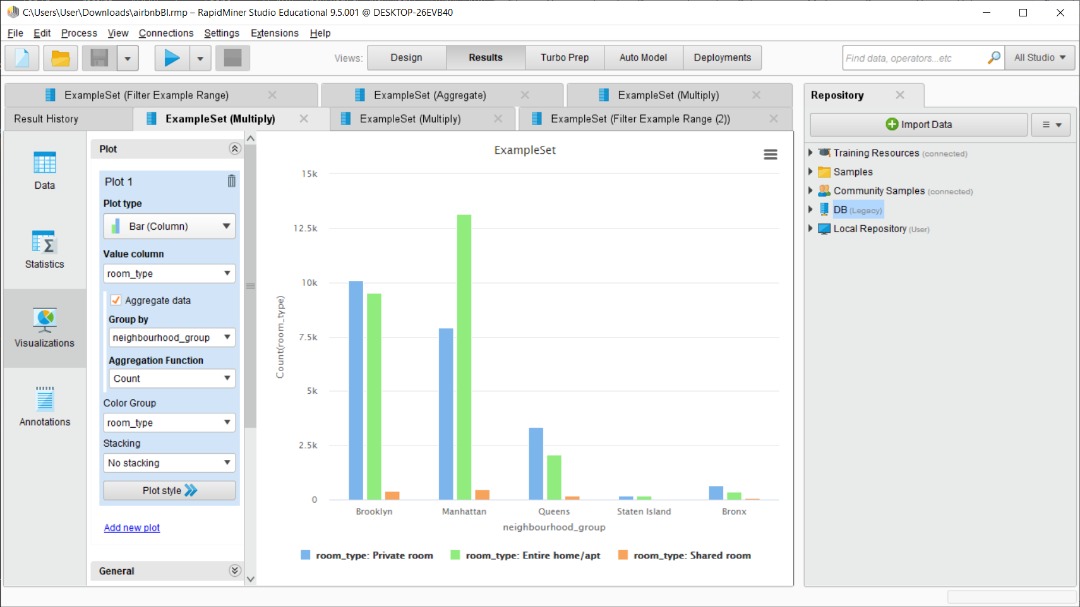


Figure 14.0 Bar chart of room type existed for each neighbourhood group

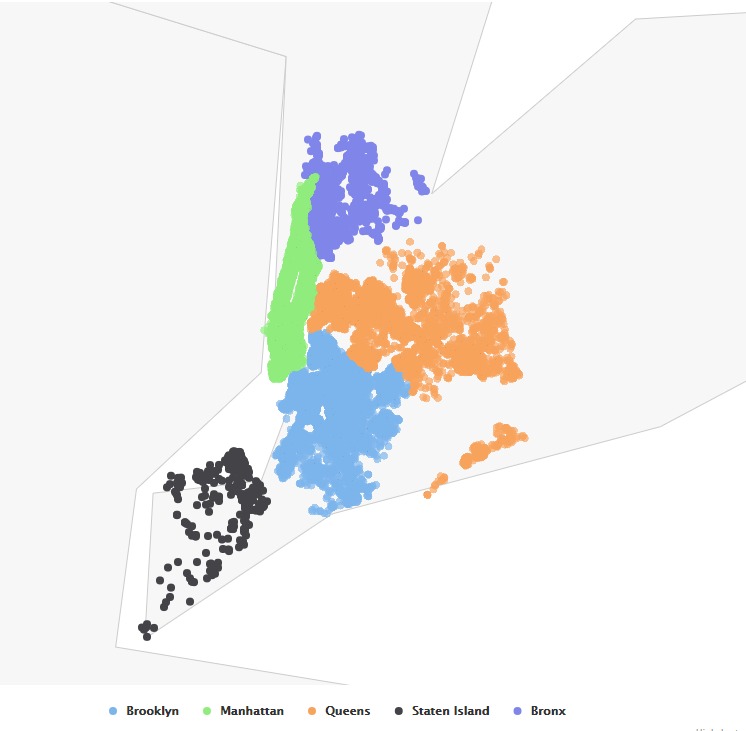


Figure 15.0 Point Mapping for room types

Conclusion:

All the room type existed for each of the neighbourhood group are plotted on the bar chart and the point map.

The total room type for Brooklyn is 20027, private room type is 10090, entire home/apt room type is 9526 and shared room type is 411.

The total room type for Manhattan is 21573, private room type is 7933, entire home/apt room type is 13163 and shared room type is 477.

The total room type for Queens is 5624, private room type is 3345, entire home/apt room type is 2082 and shared room type is 197.

The total room type for Staten Island is 366, private room type is 185, entire home/apt room type is 172 and shared room type is 9.

The total room type for Staten Island is 1074, private room type is 638, entire home/apt room type is 376 and shared room type is 60.

### 3.1.6 Does the room types of a specific neighbourhood group affect the amount of price?

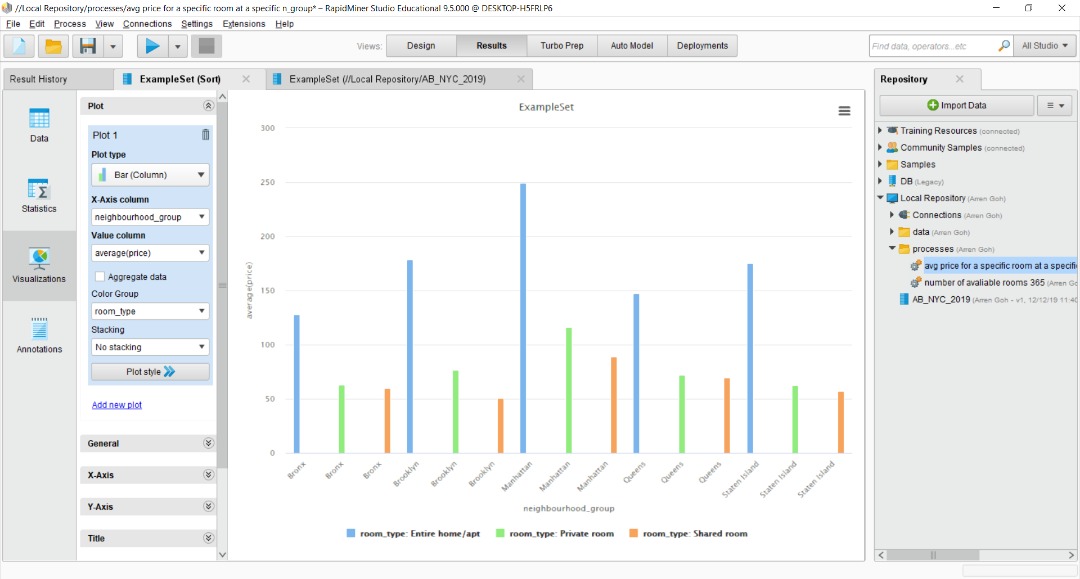


Figure 16.0 Bar chart of the relationship between room types of a specific neighbourhood group and the price

Conclusion:

Yes, different neighbourhood group affect the amount of price. Manhattan has the highest rate of price among all the other neighbourhood groups because the standard of living is high in Manhattan. Bronx has the lowest rate of price among all the other neighbourhood group because the standard of living is low in Bronx.

### 3.1.7 Predict the next room type’s price by calculating the average price of it?

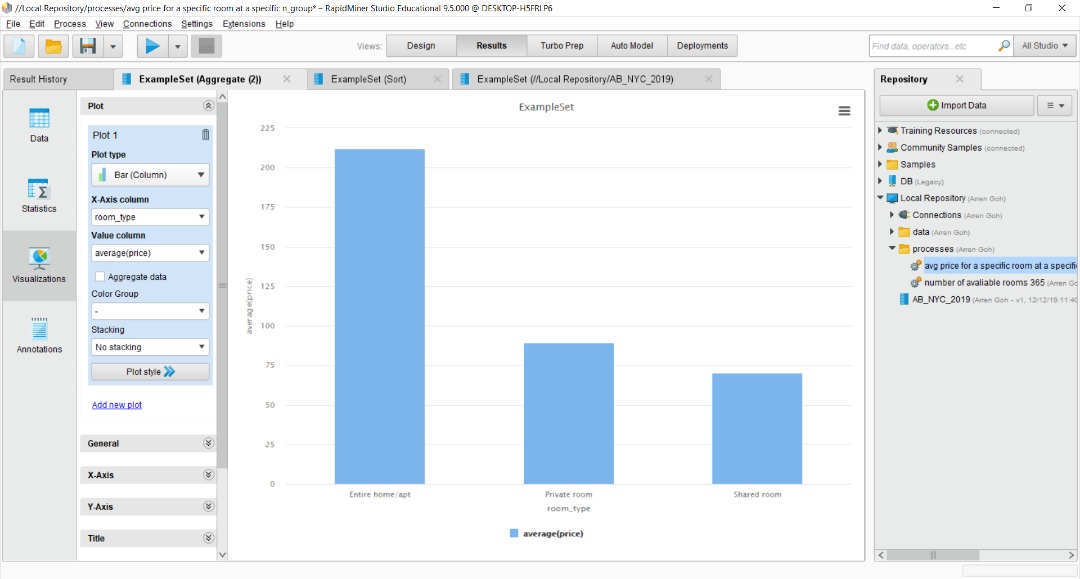


Figure 17.0 Bar chart of the average price of each room type

Conclusion:

The average price of each room types are plotted on the bar chart. The average price of entire home/apt room type is USD 211.93 hence, the next entire home/apt room type should be around USD 211.93. The average price of private room type is USD 89.44 hence, the next private room type should be around USD 89.44. The average price of shared room type is USD 70.22 hence, the next shared room type should be around USD 70.22.

### 3.1.8 Predict the each of the room type’s price for each neighbourhood\_group.



Figure 18.0 Bar Chart of Neighbourhood for each room type’s price

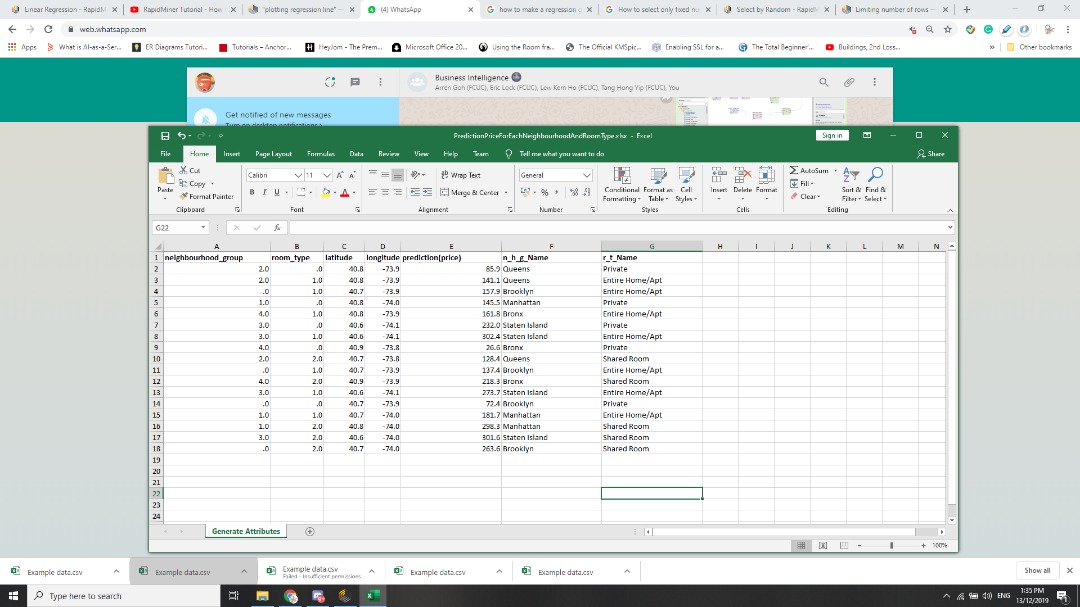


Figure 19.0 Table of Neighbourhood for each room type’s price from excel file

Conclusion:

Based on figure 16 and 17, the prediction for the next price for each of the room type in each neighbourhood from the training data set using linear regression in Rapid Miner. For instance, the predicted price for private room, entire house/apt and shared room in Queens are USD 85.9, USD 141.1 and USD 128.4.

References:

[1]"AIRBNB BUSINESS MODEL | How does Airbnb make revenue? - Vizologi | rethinking business model design", Vizologi | rethinking business model design, 2018. Accessed on: 12- Dec- 2019 [Online]. Available: <https://vizologi.com/airbnb-business-model-how-do-airbnb-works/>.

Github: <https://github.com/hongyiptang/BI-Group>