**Analysis of New York City Condo Evaluation for Fiscal Year 2011-2012**

**Abstract:** The team analyzed data from the New York City condo evaluations for the fiscal year 2011 and 2012 project. Analysis on the variables included in the dataset for each Boro was carried out in order to determine the averages of each variable. For this paper, we attempt to use this dataset to gain an insight into our each borough in New York City compares to each other as it relates to prices, values, size, and new buildings construction.

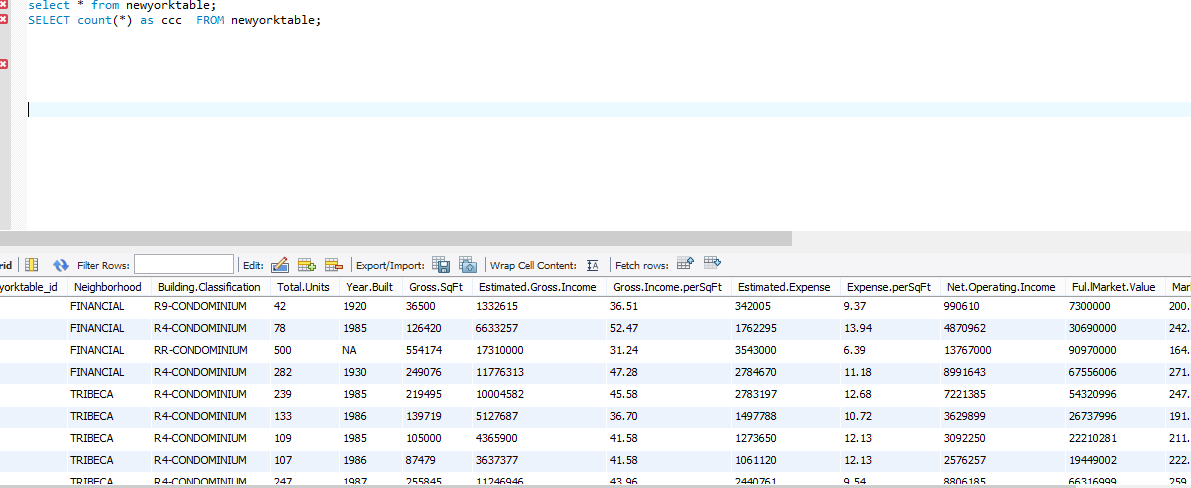
1. **Introduction**

This paper explores the structure of the New York City housing market by analyzing the various variables included in the dataset. Due to the ever growing population in New York City, prices of properties has been on the increase, the average price of condos has surpassed its last peak before the 2008 recession hit. For this paper, we explore the various aggregate functions to help us determine the questions to our answers. For each unique variable in our dataset, we want to know which of the Boros have the maximum number, minimum number, and also find out the average number for these Boros. The NYC Open Data is a 2011 project by New York City that has made available troves of data generated by public agencies about the city. The data is updated as new information comes in, and is, provided in a machine-readable format for analysts. This dataset includes information on housing data split by boroughs. Our dataset consists of 14 variables, 2626 records and 276 KB of data.

1. **Data Analysis**

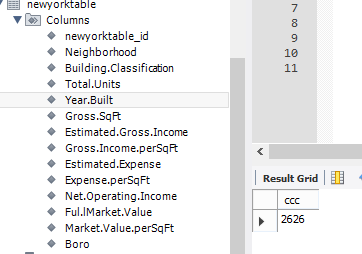
The dataset used was imported from the web into the MySQL server. The dataset was stored in a csv format. Fig 1.0 shows the dataset records

Fig 1.0



The dataset consist of 14 variables, and 2626 observations as seen in fig 1.1 below;

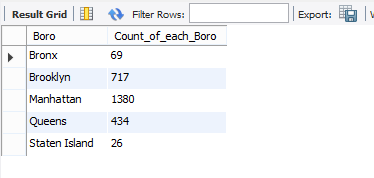
Fig 1.1



Count of Each Boro

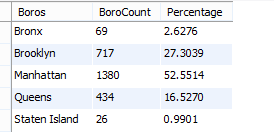
Since our analysis is based on comparing the different boros, we pulled the distinct boros from our dataset and the output shows five distinct boros located in our dataset, these datasets include Manhattan, Brooklyn, Queens, Bronx, and Staten Island. The records in the dataset were further analyzed by grouping them by Boro variable, this provided us with the number of each records under each Boro in the dataset, and from the result, and Manhattan has the highest number of records in the dataset. Fig 1.2 below shows the output.

Fig 1.2 Count of Each Boro



We further analyzed the percentage weight each Boro represented in our dataset, and we found out the Manhattan Boro consist of over half the records in our dataset. Fig 1.3 below shows the output.

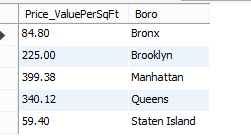
Fig 1.2 Percentage Weight vs Count (Boro)



We further did an analysis based on the Boro variable with the highest Square feet per value among the Boros, from our output, Manhattan seems to have the highest value of square feet. Fig 1.3 below attest to the fact.

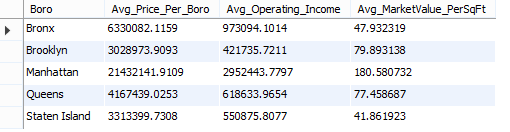
Fig 1.3

MarketValueSqFt vs Boro Analysis



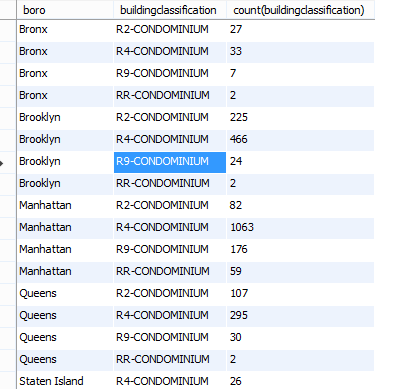
We wanted to find out which boro has the highest price per square feet, the highest operating income, and the market value per sqft. Based on our analysis, we found out that Manhattan has the highest average for Price per boro, while Brooklyn has the lowest average, for the average value per square feet, it’s not surprising that Manhattan has the highest average market value per sqft while Staten Island has the lowest average market value per sqft. Our analysis also showed that Manhattan has the highest operating income among the boros. Fig 1.4 below shows the output.

Fig 1.4 Averages Comparison using 3 variables between Boros



We analyzed the most common building types for each boro, from fig 1.5 below, in Manhattan, the R4-Condominium is the most common type, while the RR-Condominium is the least common type. Looking at the other boros, the RR-Condominium is the least common type.

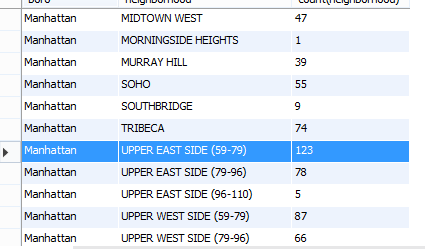
Fig 1.5 Analysis to find the most common building classification per Boro



Analysis to find the highest number of neighborhood per Boro

We analyzed the neighborhoods with the highest number of condos for each boro, for the Manhattan Boro, the Upper East Side (59-79) has the highest number of condos, for the Bronx boro, the Riverdale neighborhood has the highest number of condos. For the Brooklyn boro, the Bedford Stuyvesant has the highest number of condos, while for the Queens boro, the Flushing North has the highest number of condos and for the Staten Island Boro, New Springville has the highest number of condos. Fig 1.6 below shows the output.

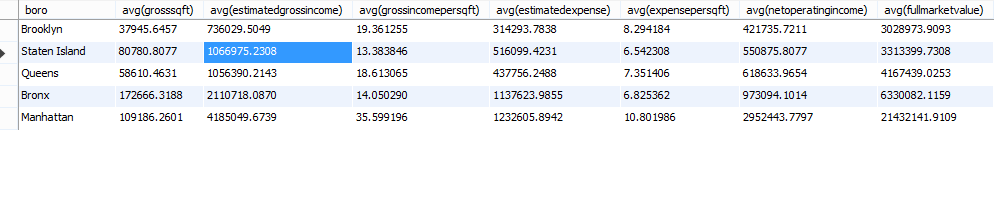
Fig 1.6 Neighborhood with the highest number for Manhattan



**3.Summary**

Based on our analysis of the dataset variables, we found out that the Manhattan Boro has the most expensive condos on the average, Manhattan also has the highest average as it relates to estimatedgrossincome, grossincomepersqft, expensepersqft, and netoperastingincome. Despite the fact that Manhattan has the most expensive condos on the average, the boro has more condos than the other boros according to our analysis. Fig 1.7 below shows the output.

Fig 1.7 Averages of Variables per Boro



**References**

<http://www.jaredlander.com/datasets/>

<https://nycopendata.socrata.com/data?cat=housing>