DEMO DATA ANALYSIS REPORT

CLIENT'S NAME: **XYZ COMPANY**

SUBJECT: **HOUSE PRICE ANALYSIS FOR XYZ COMPANY**

DATE: **XX/XX/XXXX**

**Note: Delivery of the analysis is as per the requirement agreed upon by both companies in accordance with the M.O.U. signed on XX/XX/XXXX*

Introduction:

The dataset considered, involves extraction of the information about the pricing of houses in a particular region. On critically analysing the data, it was derived that the house price varied based on multiple factors, some of these were directly related to the house price whereas some of these were inversely related. The factors obtained played a pivotal role in determining the price of the house. These factors are mentioned below.

- Average Income of residents of the city house is located in
- Average Age of Houses in same city
- Average Number of Rooms for Houses in same city
- Average Number of Bedrooms for Houses in same city
- Population of city house is located in
- Price that the house sold at
- Address for the house

Visual Representation of the data:

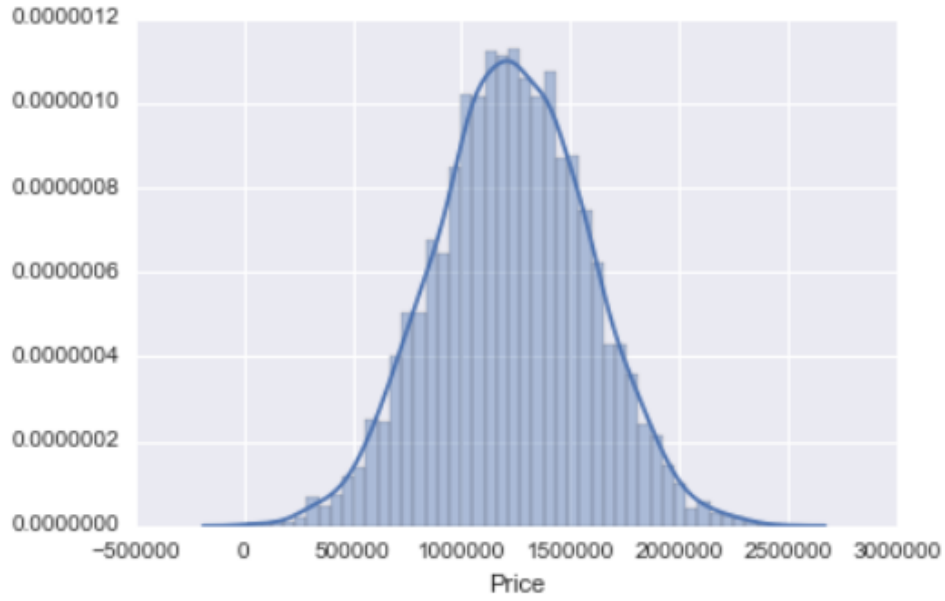
	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Address
0	79545.458574	5.682861	7.009188	4.09	23086.800503	1.059034e+06	208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
1	79248.642455	6.002900	6.730821	3.09	40173.072174	1.505891e+06	188 Johnson Views Suite 079\nLake Kathleen, CA...
2	61287.067179	5.865890	8.512727	5.13	36882.159400	1.058988e+06	9127 Elizabeth Stravenue\nDanieltown, WI 06482...
3	63345.240046	7.188236	5.586729	3.26	34310.242831	1.260617e+06	USS Barnett\nFPO AP 44820
4	59982.197226	5.040555	7.839388	4.23	26354.109472	6.309435e+05	USNS Raymond\nFPO AE 09386

Given above is not the complete data but an excerpt of the data. This data has around 5,000 records and the target feature in this report is the price column. The price in some way or the other has to be validated. This validation of price can be done by analysing the data. And this analysis can be made easy with the help of Data Visualization as it will make the traits much more conspicuous and perceivable to human eye.

Statistical information about the data:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price
count	5000.000000	5000.000000	5000.000000	5000.000000	5000.000000	5.000000e+03
mean	68583.108984	5.977222	6.987792	3.981330	36163.516039	1.232073e+06
std	10657.991214	0.991456	1.005833	1.234137	9925.650114	3.531176e+05
min	17796.631190	2.644304	3.236194	2.000000	172.610686	1.593866e+04
25%	61480.562388	5.322283	6.299250	3.140000	29403.928702	9.975771e+05
50%	68804.286404	5.970429	7.002902	4.050000	36199.406689	1.232669e+06
75%	75783.338666	6.650808	7.665871	4.490000	42861.290769	1.471210e+06
max	107701.748378	9.519088	10.759588	6.500000	69621.713378	2.469066e+06

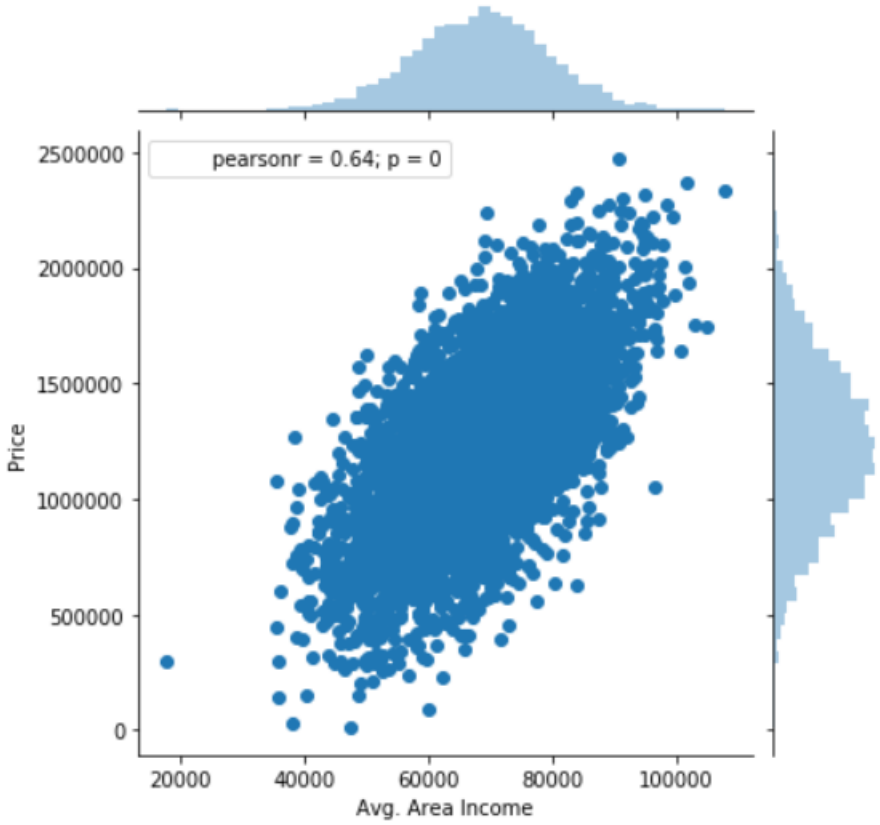
These statistics present insights about the data that one would be looking for. Maximum amount of *income group* lies around *50000-75000* and tends to prefer the house which has ***6-8* number** of ***rooms.*** Average ***number of bedrooms*** that maximum number of people are looking for is around ***3-4.*** Also, the population in the surrounding area is having an effect on the price of the house but it will need some exploratory data visualization in order to give a clear word on the same. The house that is being preferred by maximum amount of people is around ***5-6 years in age*.** These features form the features of an ideal house and the ***price*** of such house is around ***1.234456e+08-1.564323e+06***. And this can be observed via the following graph:



As observed, the price is mainly concentrated in the ***1.234456e+08-1.564323e+06*** mark and hence, ideal house will have to be paid around it. But the data just not provides this information about ideal price of the house. On some more data analysis through visual representation we have found out that there are certain traits and features that are affecting the price of the house.

Price variation with respect to features:

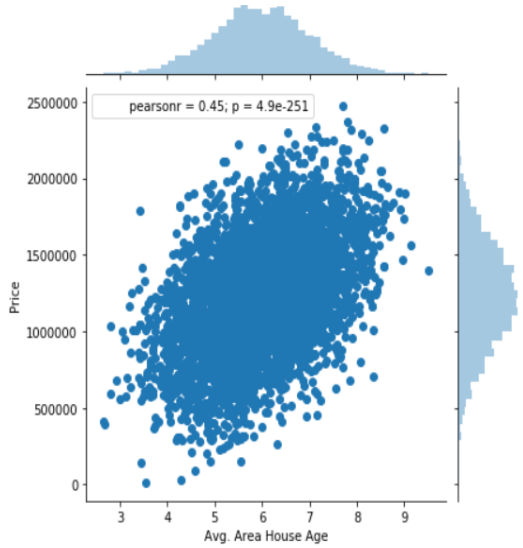
Following excerpt will enlighten the variation of the price with respect to various features that are present in the data. It is observed via the data visualization that most of the features are directly related to the house price meaning that price will increase as the related features scale up. And one such prominent feature that has its nose ahead when it comes to rising price and feature scaling up is the Average Income of the residents in the area.

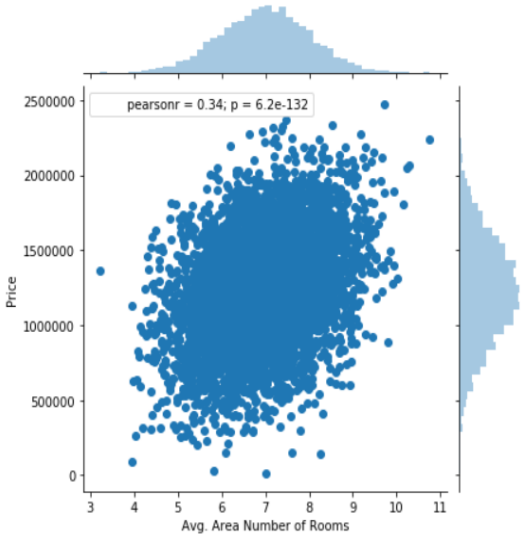


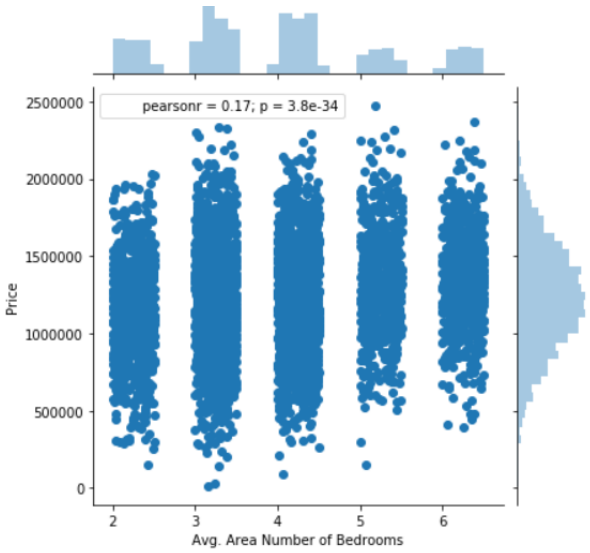
The joint plot above represents the Avg. Area Income on X-axis and Price on Y-axis. The trend observed above displays the fact that more the income, the higher price house is been bought. Also, the bar plots around the scatter plot showcase the distribution of the ***price*** which is around ***1.234456e+08-1.564323e+06 mark***. But here's a catch, the house price will increase only up to a certain mark which in this case is around ***90000 after that demand falls*** for the houses.

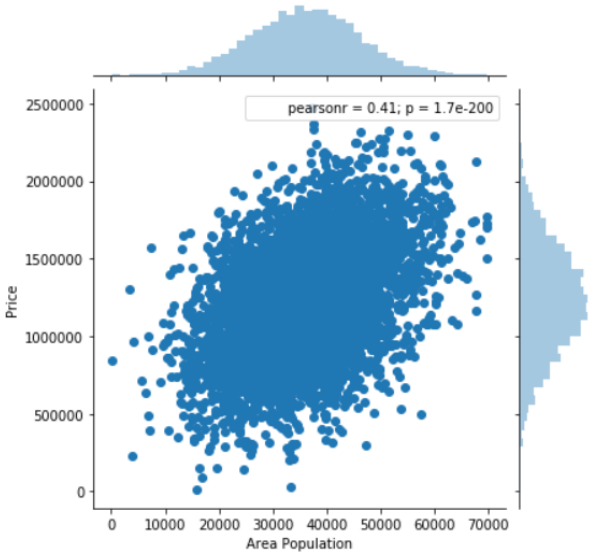
Similar trend has been observed with other features in the dataset that we have received. In most cases while exploring the data, we have kept Price as the dependent variable and rotated the independent variables along the X-axis ('Avg. Area House Age', 'Avg. Area Number of Rooms', 'Avg. Area Number of Bedrooms', 'Area Population') in order to maintain the continuity in the visualization. Just like Avg. Area Income the features mentioned before produced a similar kind of effect on the target column which is 'Price'. As these features scale up, up to a certain point the price of the house rises but after a threshold the demand for the houses falls resulting in falling of the prices of the house.

Data Visualization of other features vs Price:





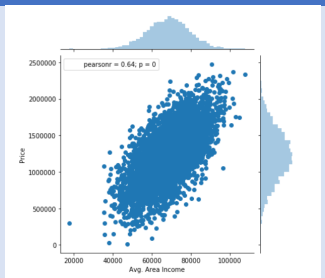
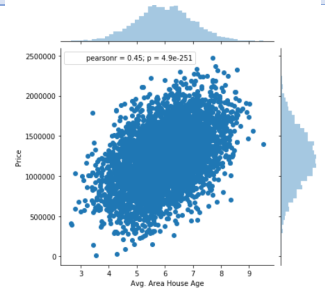
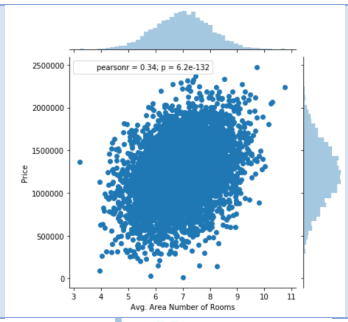
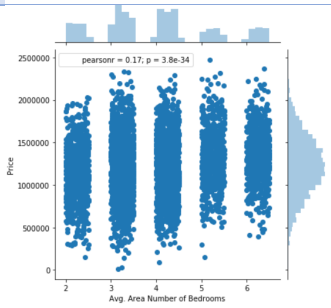
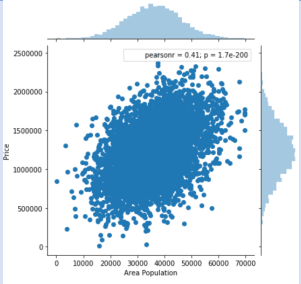




Conclusion:

The data analysis through the data visualization stands good enough to prove the fact that features are directly proportional to price but only upto a certain point. After that point the demand for the house falls resulting into less number of house being sold after that particular point.

Summary of Data Analysis:

Features	Graph vs Price	Ideal number	Effect on Price
Average Income Group		65000	Directly Proportional
Average Area House Age		6	Directly Proportional
Average Area Number of Rooms		7	Directly Proportional
Average Area Number of Bedrooms		3	Directly Proportional
Area Population		37000	Directly Proportional