Swati Sharma, xn2486  
Mihir Mane, du3425  
Deepansh Malviya, ts1799

BAN 602 Team Project

By Group 1

We have taken the data set of San Francisco housing prices containing attributes as follows:

* Prices
* Sqft
* Beds
* Baths
* Laundry
* Pets
* Housing type
* Parking
* Hood district

1. **Set of Business questions related to the data set:**

**Q1. Calculate the number of houses having the maximum number of beds.**

**Answer:** From the below-calculated values we can say that the maximum number of houses belongs to the 1-bed category.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Number of beds | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Number of houses | 106 | 370 | 322 | 134 | 43 | 11 | 3 |

**Q2. Which bath unit accounts for the maximum number of houses?**

**Answer:** From the below-calculated values we can say that 1 bath unit accounts for the maximum number of houses, which is more than half of all other houses containing more than 1 bath.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Bath counts | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |
| Number of houses | 626 | 40 | 269 | 33 | 13 | 5 | 3 |

**Q3. How many housing units have the following:**

1. **In-unit laundry**
2. **On-site laundry**
3. **No Laundry**

**Answer:** Following are the values

1. In-unit: 488
2. On-site: 370
3. No Laundry: 131

**Q4. How many house units accommodate pets?**

**Answer:** From the below data we can conclude that most houses don’t allow pets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pets | Dogs | Cats | Both | No Pets |
| Number of houses | 367 | 21 | 91 | 510 |

**Q5. Which district has the maximum number of houses?**

**Answer:** From the below-calculated values we can say that over half the houses are present in districts 8 and 9.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Districts | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| House counts | 50 | 51 | 14 | 11 | 106 | 62 | 123 | 207 | 332 | 33 |

**Q6. What is the total number of houses that have parking permits?**

**Answer:** From the below calculated we can say that 439 houses have parking permits.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parking types | Valet | Protected | Off-Street | No parking |
| House counts | 34 | 439 | 38 | 478 |

1. Summary between Price and Sqft area of houses

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Minimum | 1st Quartile | Median | 3rd Quartile | Mean | Maximum |
| Price ($) | 750 | 2650 | 3300 | 4242 | 3595 | 19000 |
| Sqft | 150.0 | 650.0 | 900.0 | 1200.0 | 976.8 | 3500.0 |

**Chart, scatter chart

Description automatically generated**

**Insights:**

* The minimum and the maximum price of the house is $750 and $19000 respectively
* We may observe a decline in value of price per square foot between the second and the third quartile (i.e. 900 – 1200 sqft).
* Based on the information above, we can conclude that the per square feet value is almost the same for all price ranges.
* According to the statistics, we may conclude that there is a linear progression between price and square foot area, meaning that a house's price rises as its square foot area increases.

1. Summary between the Bed and the Bath type:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Beds vs Baths | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |
| 0 | 106 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 349 | 12 | 9 | 0 | 0 | 0 | 0 |
| 2 | 132 | 9 | 167 | 11 | 3 | 0 | 0 |
| 3 | 35 | 14 | 65 | 13 | 5 | 2 | 0 |
| 4 | 4 | 3 | 20 | 7 | 3 | 3 | 3 |
| 5 | 0 | 2 | 7 | 1 | 1 | 0 | 0 |
| 6 | 0 | 0 | 1 | 1 | 1 | 0 | 0 |

**Chart, bar chart

Description automatically generated**

**Insights:**

* We can infer from the statistical calculations that the maximum number of house units have 1 bed and 1 bathroom.
* There are 106 housing units in total with 0 beds and 1 bathroom.
* Only houses with 4 bedrooms, which are relatively fewer in number, can accommodate more than 3.5 baths.

1. **Find the probability of a house with sqft > 1500 (Using probability Distribution)**

**Answer**: Below are the findings found as depicted in the table:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Minimum | 1st Quartile | Median | 3rd Quartile | Maximum | Mean | Standard Deviation |
| Values | 150 | 650 | 900 | 1200 | 3500 | 976.8 | 474.6289 |

Text

Description automatically generated

As can be seen from the above results, the probability of a house having area more than 1500 sqft is 13.51%.

1. **Calculate the hypothesis that the mean sqft of the houses in district 9 are equal to the mean sqft of houses in all districts.**

**Graphical user interface, text, application

Description automatically generated**

**Answer:** Here, we first formed the null and the alternate hypothesis.

* Ho: The mean sqft of houses in the 9th district is equal to the average sqft of houses in all districts, µ = 976.7654.

Ha: Mean sqft of houses in districts 9 is different from the mean sqft of houses in all districts, µ ≠ 976.7654.

* Level of significance α = 0.05.
* Value of test statistic: t = 0.26697.
* P-value = 0.7897.
* Since p-value > 0.05, the mean sqft of houses in district 9 does not fall under the rejection region. Thus, we cannot reject the null hypothesis. In conclusion, we are 95% confident that the mean value of houses in district 9 is equal to the mean value of houses in all the districts.