**Name:** Anushka Gupta

**SRN:** PES2UG20CS060

**Semester:** 5

**Subject:** Tableau Assignment 3

**Exploring Distributions Assignment**

Use HomePrices.csv dataset. The HomePrices-DataDescription.txt file describes all variables.

Open the HomePrices.csv text file in Tableau.

Drag HomePrices.csv sheet into the data connection canvas.

Create a bar chart showing the distribution of building types:

* Click on Sheet 1
* Double-click on Number of Records (or drag onto rows shelf). Note that the aggregation Sum(Number of Records) is placed on the Rows Shelf
* Drag Bldg Type to the Columns Shelf

**Question 1:** What have you learned about the distribution of building types? Which Building Type has the most homes for sale? The least?

**Answer: 1Fam** is the most purchased type of house.

**2fmCon** is the least bought type of house.

Create a histogram of Sale Price:

* Click on Sheet 2
* Double-click on Sale Price (or drag onto rows shelf)
* Select histogram in ShowMe

**Question 2**: What field did Tableau create to make the histogram? Why?

**Answer**: Naturally Tableau chose *Count of Sale Price* to develop the histogram so that the analyst can identify the number of families buying a particular house in given price range

**Question 3:** What is the shape of the histogram? Will the mean or median have a higher value? Which value (mean or median) would be a better measure for the center of the distribution?

**Answer:** The histogram is **Right-skewed.** In such histograms, the mean lies towards the right of the median, indicating that the mean will be **greater** than the median.

The **Median** would better measure for the center of the distribution.

Create a boxplot of Sale Price:

* Create new sheet
* Double-click on Sale Price (or drag onto rows shelf)
* Disaggregate measures (de-select “aggregate measures” in Analysis menu)
* Select box plot from ShowMe menu
* Drag Bldg Type and House Style to Tooltip on the Marks card

**Question 4:** Are there outliers present? If yes, are outliers above or below the center of the distribution?

**Answer:** Yes, there are outliers present. The outliers are present **Above** the distribution

**Question 5:** What house style has the highest sale price?

**Answer:** **2Story** House Style has the highest sale price, at 7,45,000.

Create a set of box plots of Sale Price for Building Type:

* Duplicate the Boxplot of Sale Price Sheet (right-click tab and select duplicate from the dropdown menu)
* Drag Bldg Type to Columns shelf

**Question 6:** Which building type has the lowest median Sale Price?

**Answer:** **2fmCon** (1,27,500)

**Question 7:** Which building type does not have any outliers in Sale Price?

**Answer:** **Twnhs**

**Question 8:** Which distribution has the largest spread? The smallest spread?

**Answer:** Largest Spread => 1Fam (3,19,100)

Smallest Spread => Duplex (97,000)

**Question 9:** If you were a real estate agent working with a new client, how would this graph be helpful to you? Based upon this view, what other questions do you want to ask*? (hint: Look at the other fields in dimensions for possible ideas. Remember that you can add any dimension to the columns shelf)*

**Answer***:*

* 1Fam being is the most bought building which happens to have the highest range amongst all.
* Asking the new client about their

1. Budget range
2. Building requirements
3. House style

can ease out the process of house-selection just by looking at the box plot and it’s corresponding parameters especially the Upper Hinge, Lower Hinge and Median.

* Asking the client about the minimum condition expectations can narrow down towards the optimal purchase. More the information, better for both the parties.
* Also asking about any particular features and facility that the client wishes to have could be of great help.