# Coding Exercises – Set 1

There are two datasets that have been provided

- SalesData.xlsx
- diamonds.csv

## Questions 1 – 6: Utilize the sales data set

Data Description: The sales data contains transactional sales information for each sales person. It also contains the date of sales, item sold, price of each item, sales amount, region and their corresponding manager information.

1. Find the least amount sale that was done for each item.
2. Compute the total sales for each year and region across all items
3. Create new column 'days_diff' with number of days difference between a reference date passed and each order date
4. Create a dataframe with two columns: 'manager', 'list_of_salesmen'. Column 'manager' will contain the unique managers present and column 'list_of_salesmen' will contain an array of all salesmen under each manager.
5. For all regions find number of salesman and total sales. Return as a dataframe with three columns - Region, salesmen_count and total_sales
6. Create a dataframe with total sales as percentage for each manager. Dataframe to contain manager and percent_sales

## Questions 7 - 12 Utilize the diamonds data set

Data Description: The diamonds data set contains the various dimensions and information for each diamond.

1. Count the duplicate rows of diamonds dataframe.
2. Drop rows in case of missing values in carat and cut columns.
3. Subset the dataframe with only numeric columns.

4. Compute volume as (x*y*z) when depth is greater than 60. In case of depth less than 60 default volume to 8.
5. Impute missing price values with mean.
6. In diamonds data set Using the volume calculated above, create bins that have equal population within them. Generate a report that contains cross tab between bins and cut. Represent the number under each cell as a percentage of total.