Project 1 Solution:-

Q1. Observe if there are missing values in the dataset and report the number of missing values if present. Replace the missing data with suitable values. Support your answers with appropriate inferences. (Hint: Use median value to replace missing values in numeric columns and use mode value to replace missing values in categorical columns. Removing the missing value rows is not recommended.) (10 points)

Answer: -

Part 1-

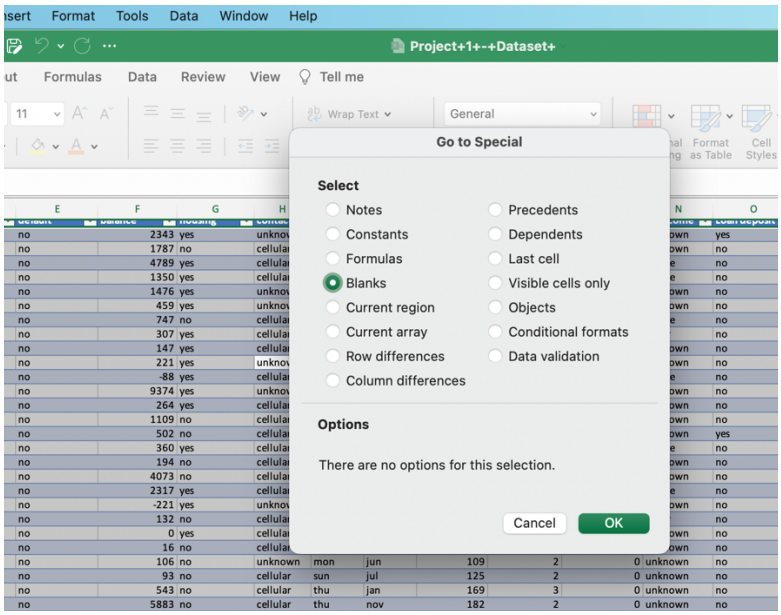
I Used 2 approaches to find and confirm the number of blank cells:

1st Approach: -

Used the Formula **=COUNTBLANK(Table1)** to find the number of blank cells. The result was that there were total 17 blank cells.

2nd Approach: -

Select the entire table and then go to **home ribbon > find & select > Go to special > blanks**:



And then filled color to highlight blank cells. Counted the highlighted cells to verify number of blank cells. This way also, the result was same i.e. there were 17 blank cells in column balance.

Part 2-

Since balance is a numeric column, I calculated the median of the balance (459), and filled all the blank cells with value 459.

I used 2 ways to calculate the median.

1st-

Used formula **= MEDIAN(Table2[balance])**

2nd-

Used descriptive analysis package to find the median of the balance column.

	no			
'n	no			
'n	no			
	no			
	no			
'n	no			
'n	no	459		
	no			
'n	no			
'n	no		*Column1*	
'n	no			
'n	yes		Mean	1477.190428
	no		Standard Error	88.43752675
'n	no		Median	459
'n	no		Mode	0
	no		Standard Deviation	2771.356095
'n	no		Sample Variance	7680414.607
	no		Kurtosis	18.60595268
'n	no		Skewness	3.666405296
'n	no		Range	28645
'n	no		Minimum	-1680
'n	no		Maximum	26965
'n	no		Sum	1450601
'n	no		Count	982
'n	no			
	no			
'n	no			
'n	no			
'n	yes			
'n	no			

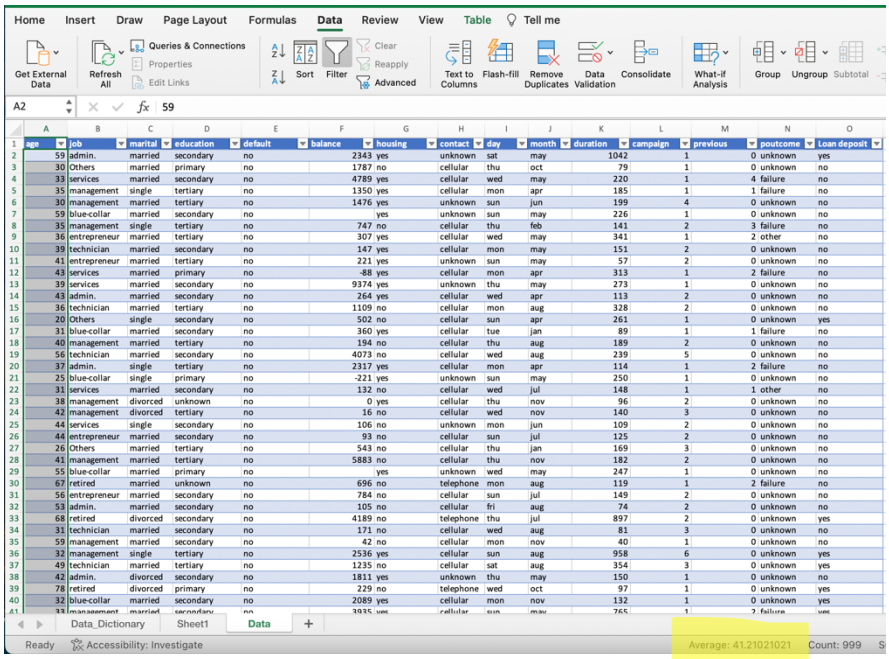
Q2. What is the average age of the customer base? What percentage of the customers in the given data set are less than 45 years old? Support your answer with appropriate inferences. (5 points)

Answer:-

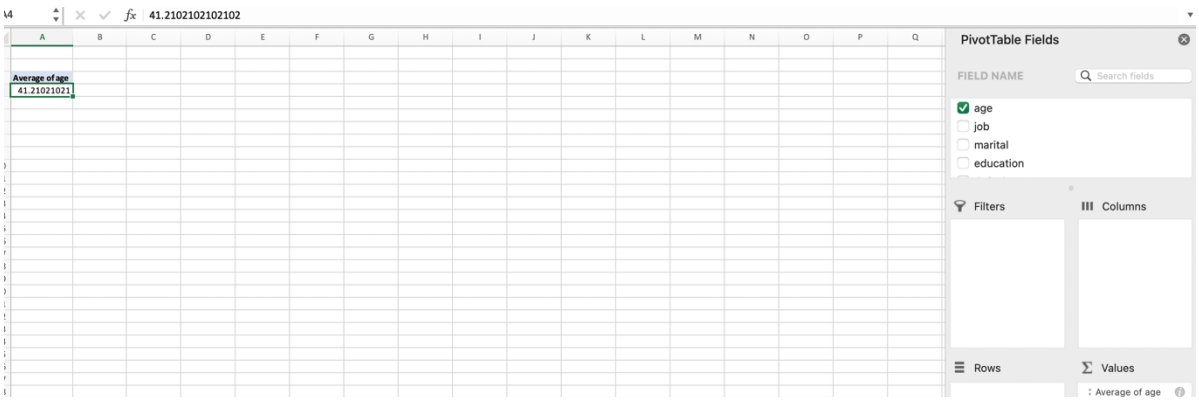
Part:1 – 41.2 years

There are multiple approached to find the average of the numeric column (“age” in this case). Below are the two examples that I used:

1. I selected the entire age column which automatically highlights the average value of the column at the bottom right of the screen. Here average age is **41.2 years**.

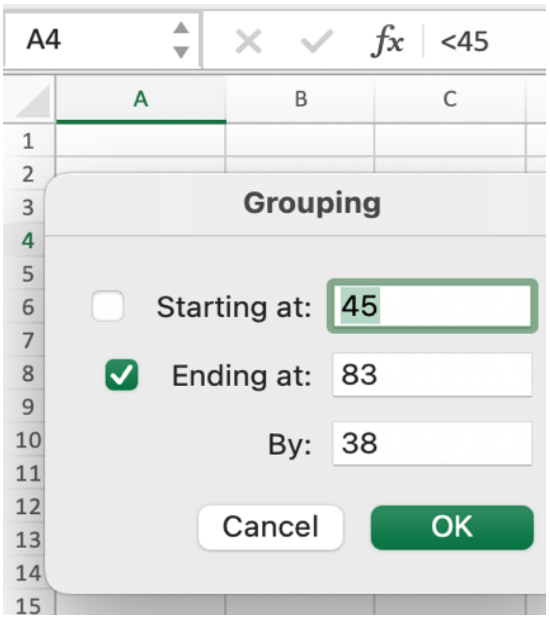


2. Another solution is to insert the pivot table and calculate the average age from there. Here as well, the answer is 41.2 years.

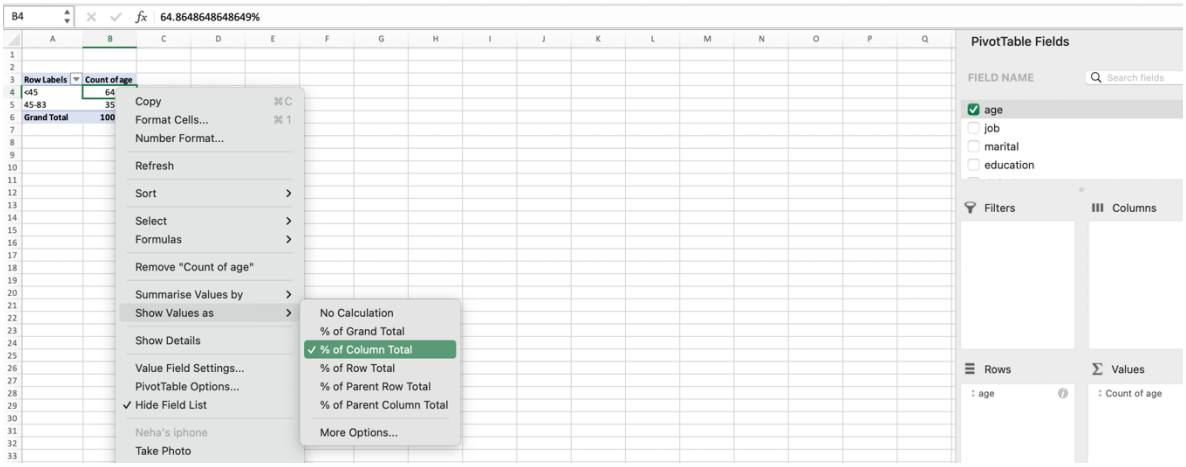


Part 2:-

Percentage of the customers in the given data set less than 45 years old is **64.86%.** To calculate the % of customers below age **45** I used the pivot table and grouped it by the age as given below



Finally displayed the count in percentage using **Show value as > % of Column Total**



Row Labels	Count of age
<45	64.86%
45-83	35.14%
Grand Total	**100.00%**

Inference: More % of people with age less than 45 are customer of bank.

Q3. What are the minimum and maximum account balances of the customer base given in the dataset? Support your answer with an appropriate screenshot for the formulas used. (5 points)

Answer:-

Minimum Account Balance = **-1680**

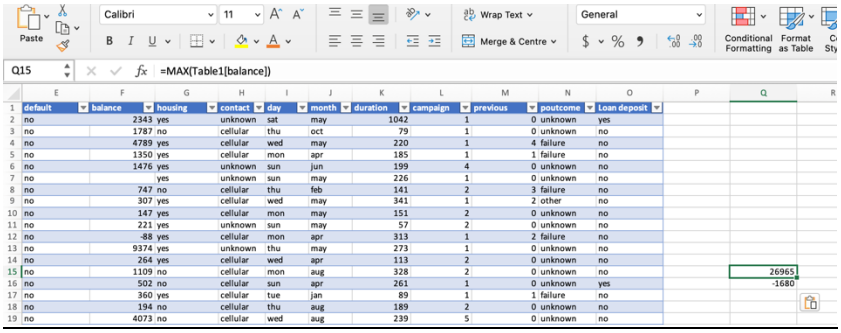
Maximum Account Balance = **26965**

There are multiple ways to find the minimum and maximum value in numerical column (balance). Like pivot table, descriptive analysis, Formulas in excel sheet.

Below explained is how I calculated using formulas from excel sheet.

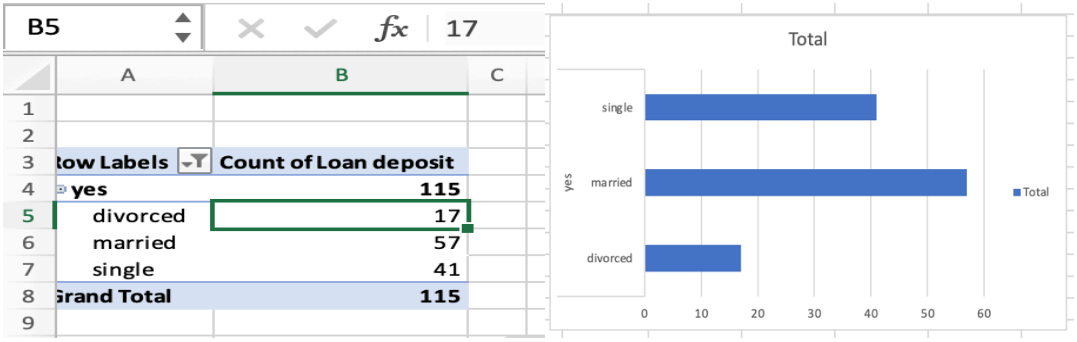
To find maximum from balance column:- **=MAX(Table1[balance])**

To find minimum from balance column:- **=MIN(Table1[balance])**



Q4. A marketing specialist claims that among the customer categories in the marital column that made a loan deposit (married, single, and divorced) people who are married made more loan deposits when compared to other categories in this column. With the help of visual analysis, approve or disapprove the claim? Support your answer with appropriate inferences. (10 points).

Answer:-



As observed from the pivot table, marketing specialist made the right claim that people who are married made more loan deposit when compared to other categories in the same column.

We can infer from the graph that of the total of 115 people who made loan deposits, **57** were married, marking their categories as the highest loan depositors, closely followed by singles where **41** made the loan deposits. There are only **17** divorced customers who made the loan deposits.

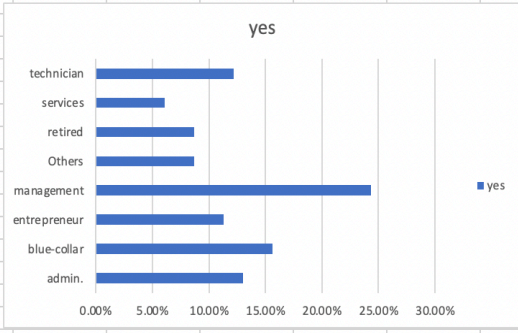
Q5. What percentage of customers have made a loan deposit with us? Among customers who made a loan deposit with us, which job category has the highest percentage of customers? Use visual analysis to gain insights and provide inferences. (10 points)

Answers:-

Row Labels	Count of Loan deposit
no	88.49%
yes	11.51%
Grand Total	100.00%

Total 11.51% of the customers have made a loan deposit with the bank as depicted in the pivot table above.

Count of Loan deposit	Column Labels	
Row Labels	yes	Grand Total
admin.	13.04%	13.04%
blue-collar	15.65%	15.65%
entrepreneur	11.30%	11.30%
management	24.35%	24.35%
Others	8.70%	8.70%
retired	8.70%	8.70%
services	6.09%	6.09%
technician	12.17%	12.17%
Grand Total	100.00%	100.00%



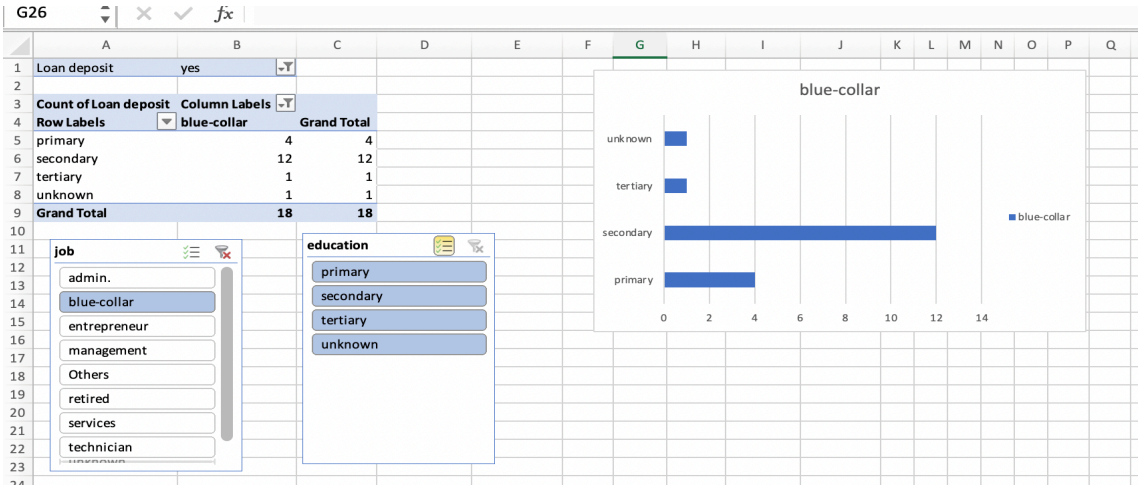
Maximum number of customers who made the loan deposits are from management category. Of the total customers who made the loan deposit management marks **24.35%,** followed by blue-collar, but still there is huge difference of almost **9%** between these 2 job categories who made loan deposit. With these findings, we can safely decide that bank should launch a campaign targeting customers from management job category who are also married.

Q6. A marketing executive claims that among blue-collar customers who made a loan deposit, customers with a secondary level of education are more likely to make a loan deposit than customers with other levels of education. With the help of visual analysis can you prove or disprove the above statement? Support your answer with appropriate inferences. (10 points)

Answers:-

When observed with the help of Pivot Table and visualization bar graph, the claim made by marketing executive that among blue-collar customers who made a loan deposit, customers with a secondary level of education are more likely to make a loan deposit than customers with other levels of education was found to be true.

Blue collar customers with secondary level of education are the highest depositors of loan by huge margin. Of the total number of blue collared customer who made loan deposit, 12% are having secondary level of education followed by 4% customers who have Primary level of education. Basis above finding, we can have this set of customers as second priority for campaign.

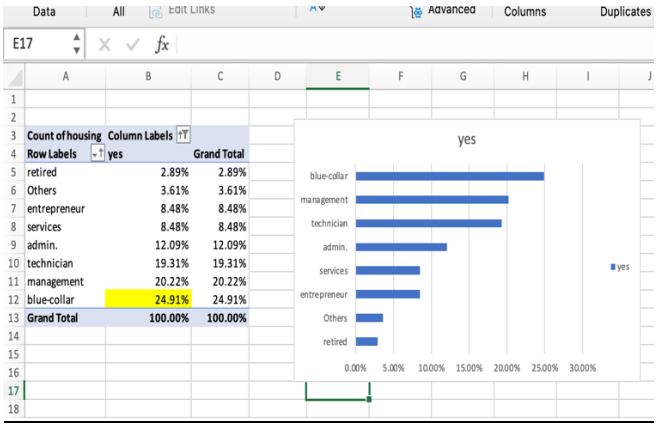


Q7. The following question will have 3 subparts. Support your answers with appropriate inferences and visualizations.

A) Under the job column, what category of people have taken the highest number of home loans? Express figures in percentages. (3 points)

Answers:-

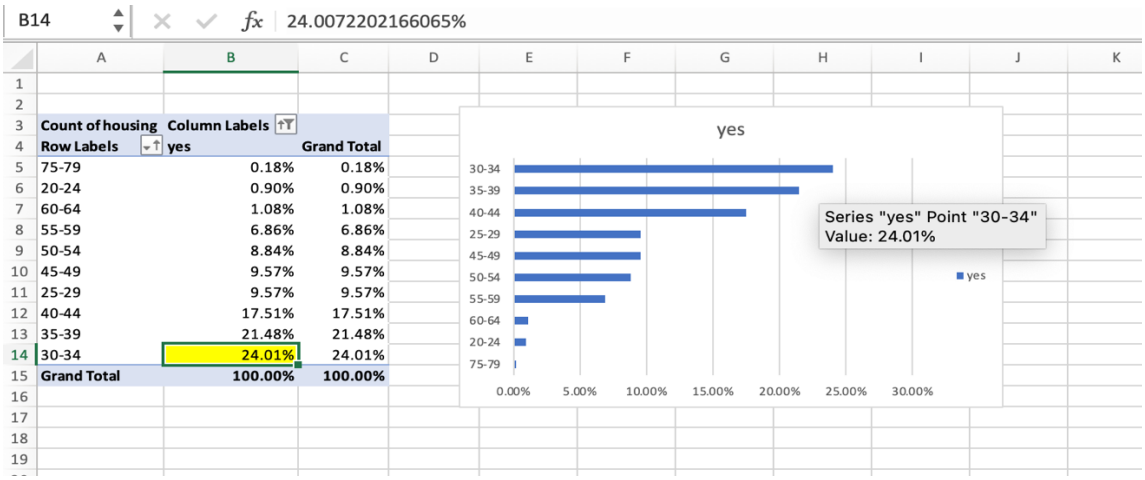
After filtering out the yes category (who have taken housing Loan) from the housing variable and observing Pivot table and Horizontal bar graph, we inferred **24.91%** of the total number of people who have taken the housing Loan, belong to Blue-collar job category which makes blue collar job category amongst all other job categories as the highest borrower of housing loans.



B) Which age group of people has the highest percentage of housing loans (Group age by the difference of 5)? (2 points)

Answers:-

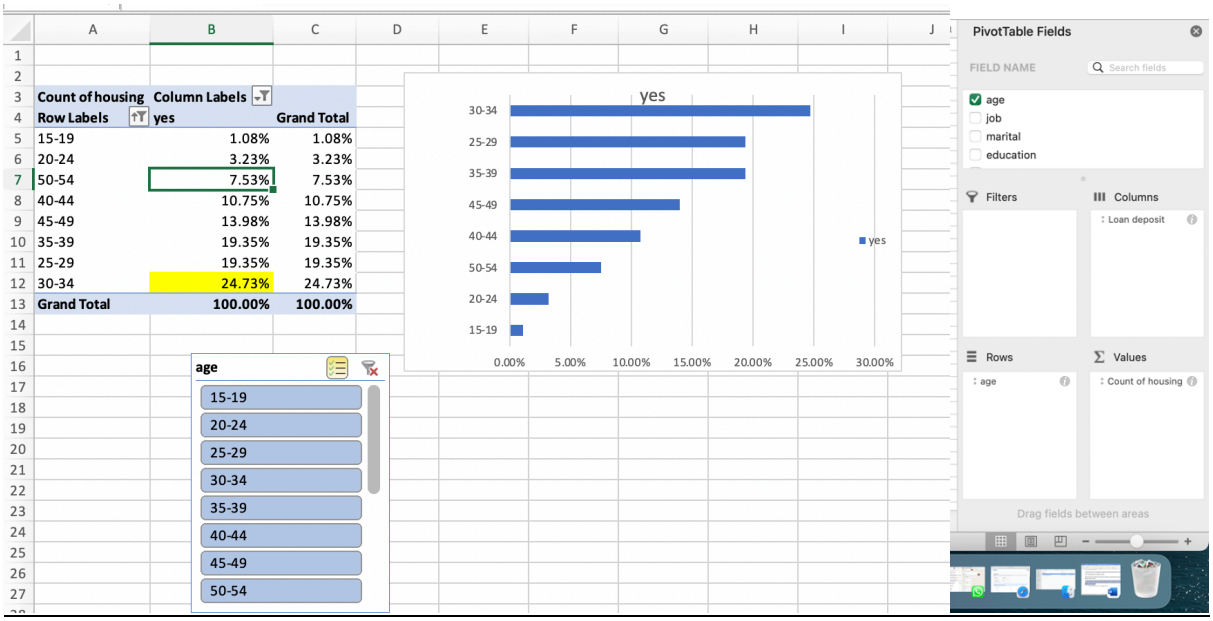
After filtering out the yes category (who have taken housing Loan) from the housing variable and observing Pivot table along with Horizontal bar graph, we inferred that **24.01%** of the total number of the people who have taken the house loan fall within age group 30-34, making themselves the highest loan borrowers as compared to other age groups.



c) Which age group of people has made the highest percentage of loan deposits? Within that age group, which job category has the highest percentage of loan deposits (group age by a difference of 5)? (5 points)

Answers:-

First we found out from the pivot table and horizontal bar graph that people falling within age group 30-34 contributes to be the highest percentage of Loan depositor i.e. 24.73%.



Then we inserted slicer on Age category, choosing age group 30-34 and then analyzing it against job categories from where we inferred that management job category has the highest percentage **(24.73%)** of loan deposit.