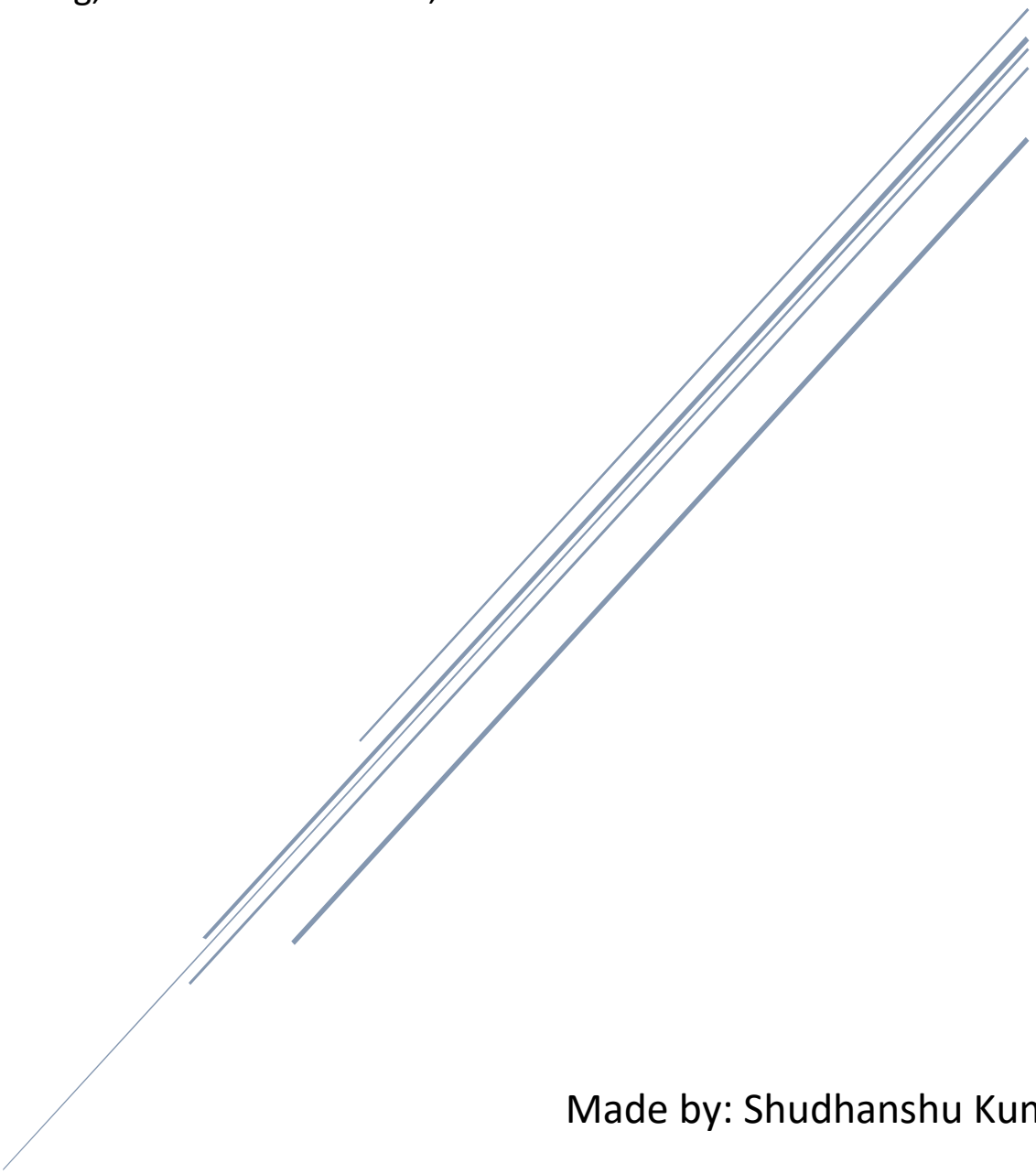
MS EXCEL PROJECT

This Project is on Bike Buyers dataset

In this Project/Case study I have explained how I performed the Data Cleaning,Created PIVOT Table,Charts and Dashboards in MS Excel



Made by: Shudhanshu Kumar

OVERVIEW OF DATASET

This dataset is taken by the Alex github page for the project purpose.

Dataset's location at the github is follow below:

https://github.com/AlexTheAnalyst/Excel-Tutorial/blob/main/Excel%20Project%20Dataset.xlsx

Dataset name-- Bike_Buyers_dataset

Column names

ID, Marital Status, Gender, Income, Children, Education, Occupation, Home Owner, Cars, Commute Distance, Region, Age, Purcahsed Bike

Dataset is about the customer who purchase the bike or not and his/her all demographic and personal details

Column name explanation:

ID: Unique ID of the customer

Marial Status: If they are Married or Single

Gender: Female or Male

Income: How much they earn

Children: How many children they have

Education: How much they have studied

Occupation: What they do

Home Owners: Are they home owners or not, Yes or No

Cars: How many cars they have

Commute Distance: How much they drive everyday

Region: In which region they live

Age: Age details

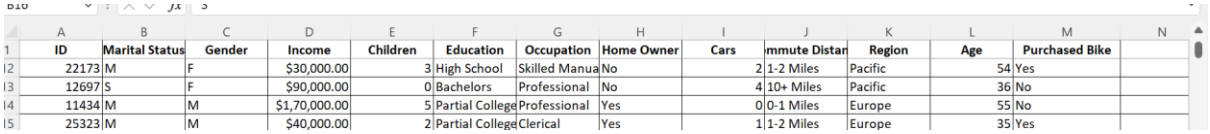
Purchased Bike- Did they purchase the bike or not, Yes or No

STEP 1: DATA CLEANING

We understood about the dataset now 1st step performing the Data Cleaning where I looked for any Duplicate record, redundant data, Null value, Missing records/values , wrong data type etc.

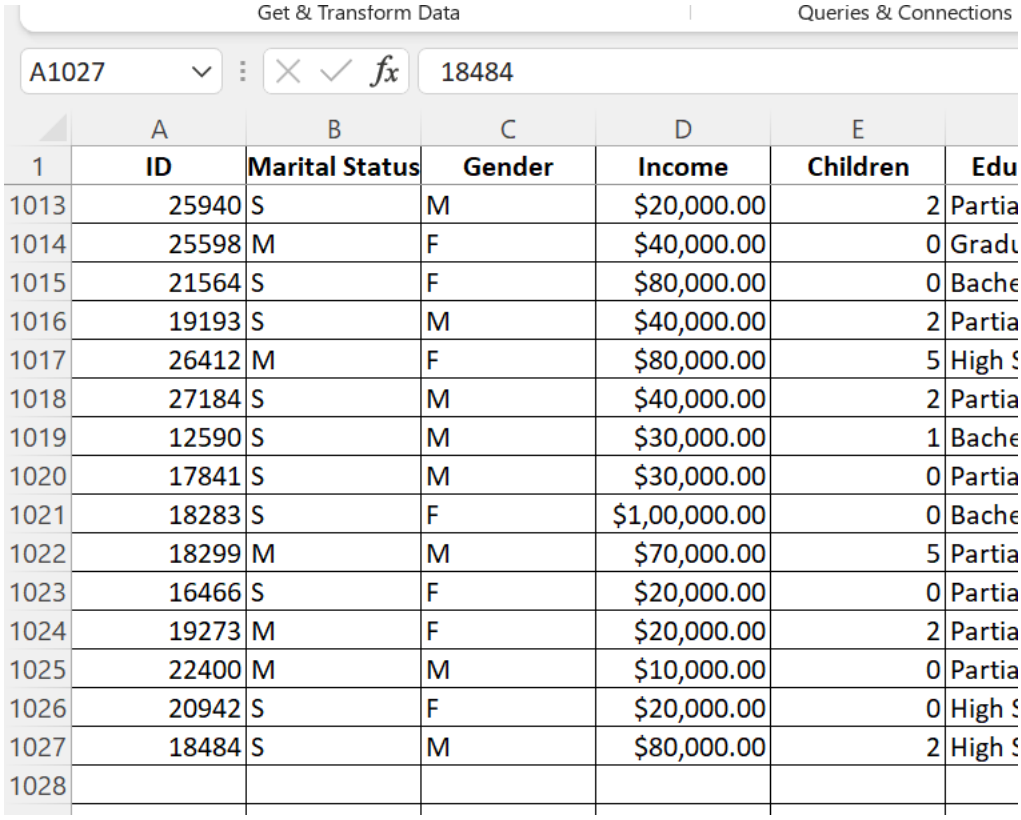
Steps followed during Data Cleaning are:

1. Freezed the top row by using Freeze pane option under View tab



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	ID	Marital Status	Gender	Income	Children	Education	Occupation	Home Owner	Cars	mmute Distar	Region	Age	Purchased Bike	
2	22173	M	F	$30,000.00	3	High School	Skilled Manua	No	2	1-2 Miles	Pacific	54	Yes	
3	12697	S	F	$90,000.00	0	Bachelors	Professional	No	4	10+ Miles	Pacific	36	No	
4	11434	M	M	$1,70,000.00	5	Partial College	Professional	Yes	0	0-1 Miles	Europe	55	No	
5	25323	M	M	$40,000.00	2	Partial College	Clerical	Yes	1	1-2 Miles	Europe	35	Yes	

2. Selected all the data and Remove duplicates under Data tab
   Before removing duplicates,we have 1027 rows

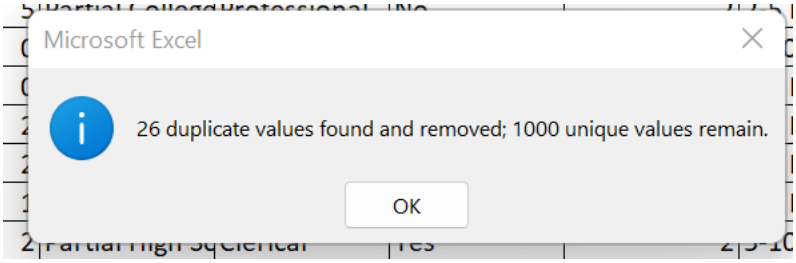


Get & Transform Data | Queries & Connections

A1027 | 18484

	A	B	C	D	E	
1	ID	Marital Status	Gender	Income	Children	Edu
1013	25940	S	M	$20,000.00	2	Partia
1014	25598	M	F	$40,000.00	0	Gradu
1015	21564	S	F	$80,000.00	0	Bache
1016	19193	S	M	$40,000.00	2	Partia
1017	26412	M	F	$80,000.00	5	High S
1018	27184	S	M	$40,000.00	2	Partia
1019	12590	S	M	$30,000.00	1	Bache
1020	17841	S	M	$30,000.00	0	Partia
1021	18283	S	F	$1,00,000.00	0	Bache
1022	18299	M	M	$70,000.00	5	Partia
1023	16466	S	F	$20,000.00	0	Partia
1024	19273	M	F	$20,000.00	2	Partia
1025	22400	M	M	$10,000.00	0	Partia
1026	20942	S	F	$20,000.00	0	High S
1027	18484	S	M	$80,000.00	2	High S
1028						

26 duplicates values were removed.



1000 Unique rows in dataset remains

A1001 | 12121

	A	B	C	D	E
1	ID	Marital Status	Gender	Income	Children
987	13920	S	F	$50,000.00	4
988	23704	S	M	$40,000.00	5
989	28972	S	F	$60,000.00	3
990	22730	M	M	$70,000.00	5
991	29134	M	M	$60,000.00	4
992	14332	S	F	$30,000.00	0
993	19117	S	F	$60,000.00	1
994	22864	M	M	$90,000.00	2
995	11292	S	M	$1,50,000.00	1
996	13466	M	M	$80,000.00	5
997	23731	M	M	$60,000.00	2
998	28672	S	M	$70,000.00	4
999	11809	M	M	$60,000.00	2
1000	19664	S	M	$1,00,000.00	3
1001	12121	S	M	$60,000.00	3

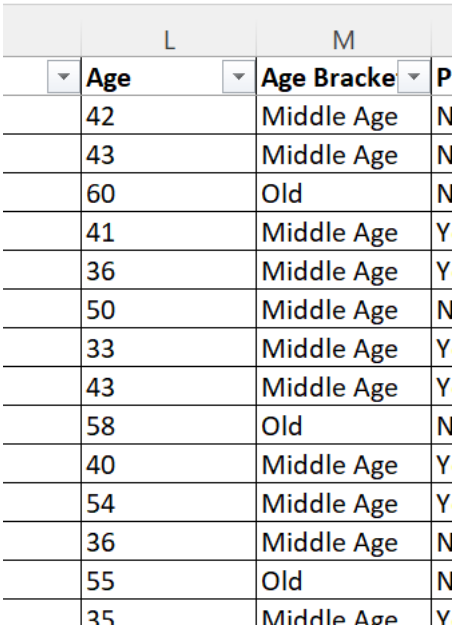
3. Replaced the Marital status Column's value from
   M to Married
   S to Single, for better understanding
   Same for Gender column from M to Male and F to Female

	A	B	C
1	ID	Marital Stat	Gender
2	12496	M	F
3	24107	M	M
4	14177	M	M
5	24381	S	M
6	25597	S	M
7	13507	M	F

4. Created one more column Age Bracket and used the Nested IF statement to categorize the age for more analysis.

=IF(L8 > 54,"Old",IF(L8>=31,"Middle Age",IF(L8<31,"Adult","Invalid")))

- If they are less than 31, Adult and if they are more than 30, else Invalid
- Greater than or equal to 31, then Middle Age
- Greater than 54, then Old



	L	M	
	Age	**Age Bracke**	**P**
	42	Middle Age	N
	43	Middle Age	N
	60	Old	N
	41	Middle Age	Y
	36	Middle Age	Y
	50	Middle Age	N
	33	Middle Age	Y
	43	Middle Age	Y
	58	Old	N
	40	Middle Age	Y
	54	Middle Age	Y
	36	Middle Age	N
	55	Old	N
	35	Middle Age	Y

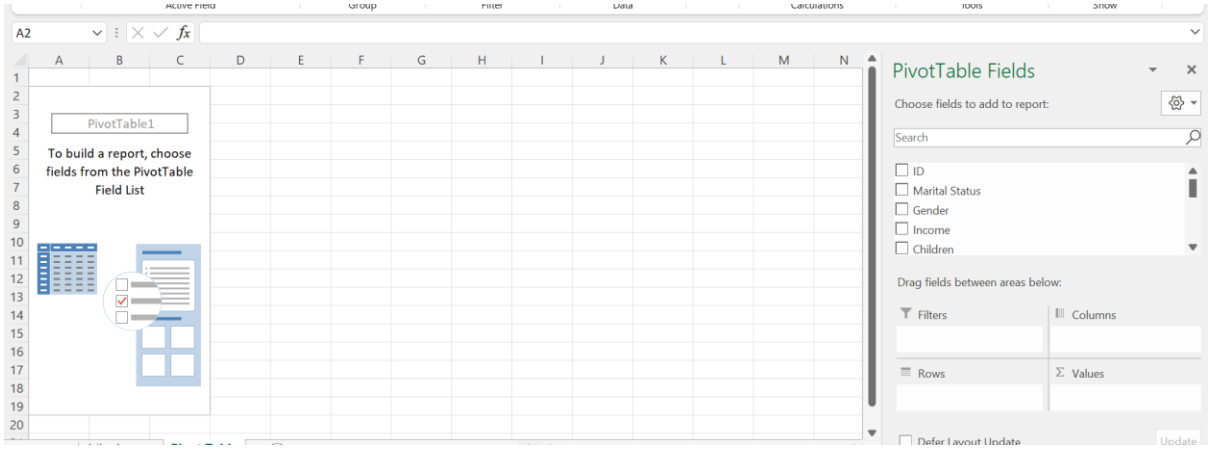
Data cleaning is completed on Raw data ,No missing values found and others all good.

STEP 2: CREATING PIVOT TABLE

Created Pivot table in Pivot table new sheet,

Selected Option:

Insert and selected the Range from bike_buyers sheet.



1.First analysis using Pivot table
What's the average salary of Male and Female customer and If they bought the bike or not?

Rows- Gender
Columns-Purchased bike
Values-Income

Average of Income	Column Labels		
Row Labels	No	Yes	Grand Total
Female	53440	55774.05858	54580.7771
Male	56208.17844	60123.96694	58062.62231
Grand Total	**54874.75915**	**57962.57796**	**56360**

From above analysis,

Female have average salary of Rs 53440 and Male have Rs 56208.17 who didn't bought the bike

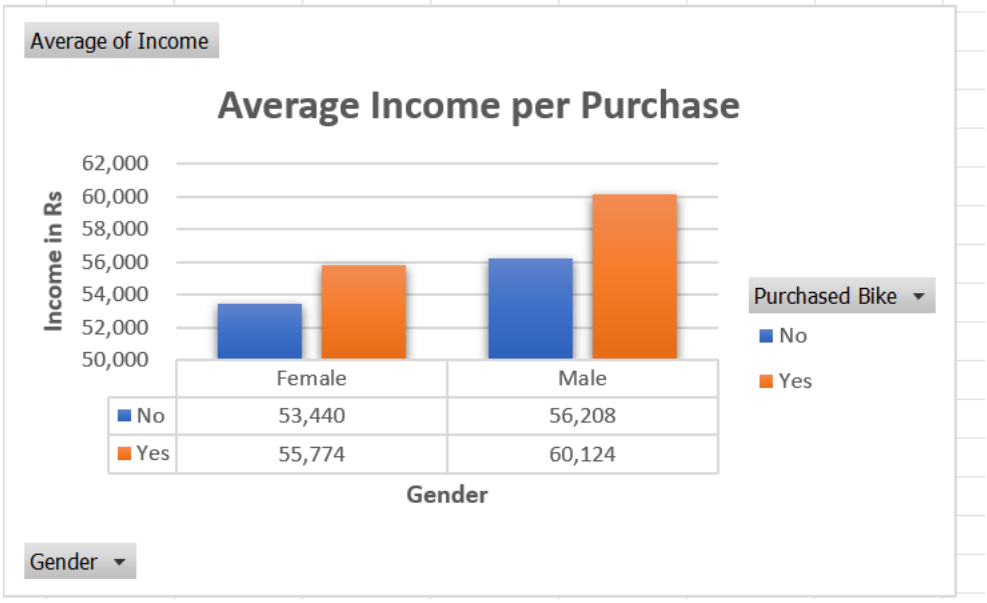
Female have average salary of Rs 54580.77 and Male have Rs 58062.62 who bought the bike

- Converted the decimal values to whole values
- Selected the cell and Right click Format cells

Average of Income	Column Labels		
Row Labels	No	Yes	Grand Total
Female	53,440	55,774	54,581
Male	56,208	60,124	58,063
Grand Total	**54,875**	**57,963**	**56,360**

- **Created Clustered Column chart**

  Insert-Recommended Chart and edited the title,axes,data labels etc.



2.Second analysis

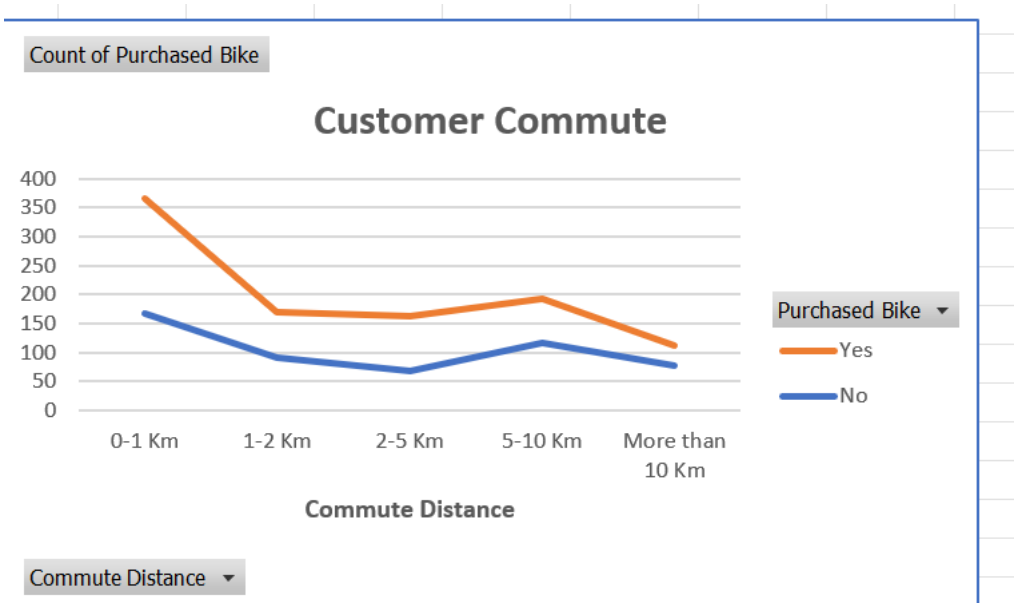
About their Salary,Why did they buy bike if they have more money or depend on the commute distance they wanted a bike to travel.

Rows-Commute Distance

Columns/Values-Purchased Bike

Count of Purchased Bike	Column Labels		
Row Labels	No	Yes	Grand Total
0-1 Km	166	200	366
1-2 Km	92	77	169
2-5 Km	67	95	162
5-10 Km	116	76	192
More than 10 Km	78	33	111
Grand Total	**519**	**481**	**1000**

Created Stacked Line chart



3.Third Anlaysis

From age bracket column,Did they purchase the bike or not?

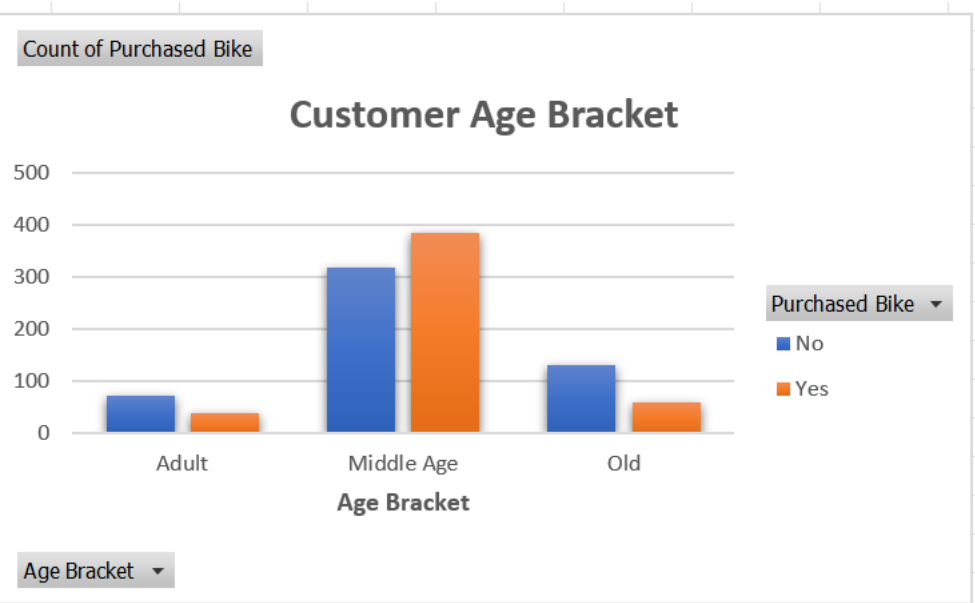
How many bike was purchased from each age group?

Row-Age Bracket

Column/Values-Purchased Bike

Count of Purchased	Column Labels		
Row Labels	No	Yes	Grand Total
Adult	71	39	110
Middle Age	318	383	701
Old	130	59	189
Grand Total	**519**	**481**	**1000**

Created Column chart and It shows Middle age bought more number of Bike



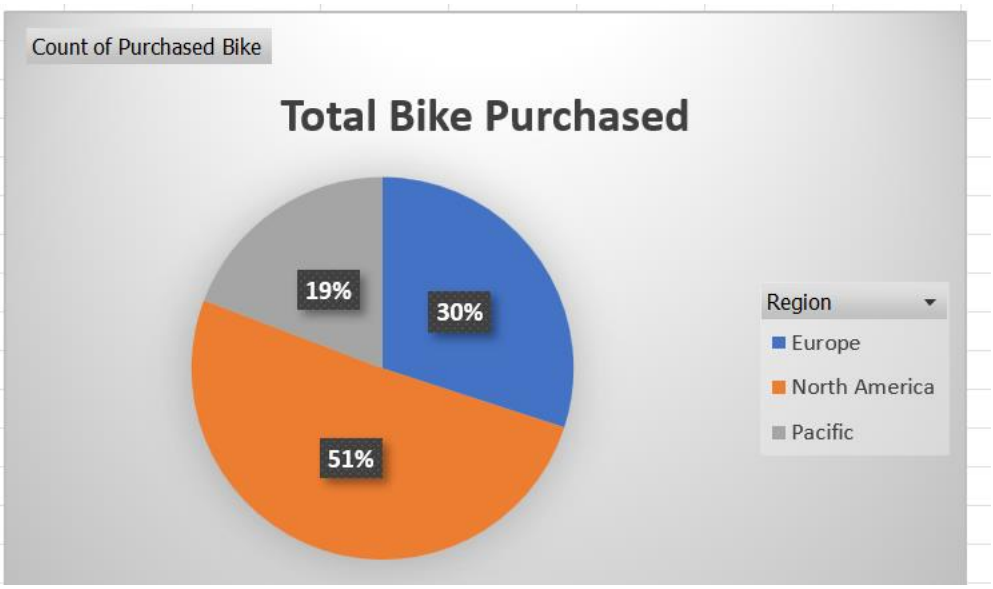
4.Fourth Analysis

How many bike were bought from each region?

Row Labels	Count of Purchased Bike
Europe	300
North America	508
Pacific	192
Grand Total	**1000**

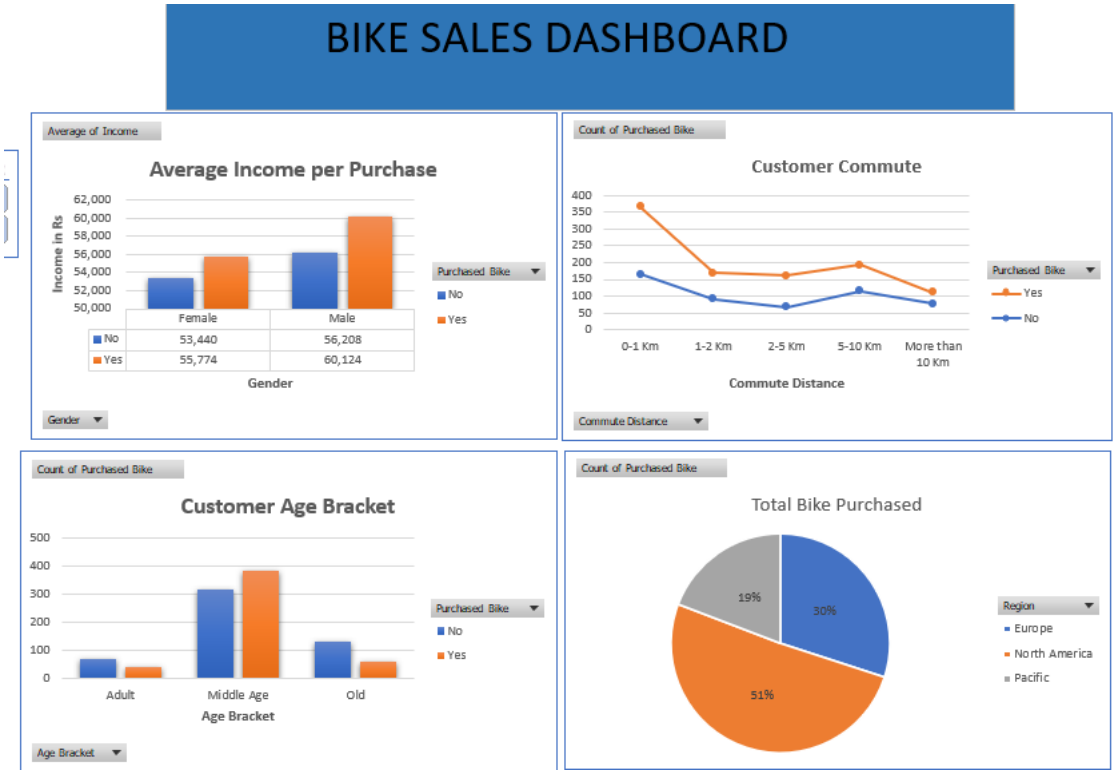
From North America Region ,more bike were purchased by the customer.

Created Pie chart

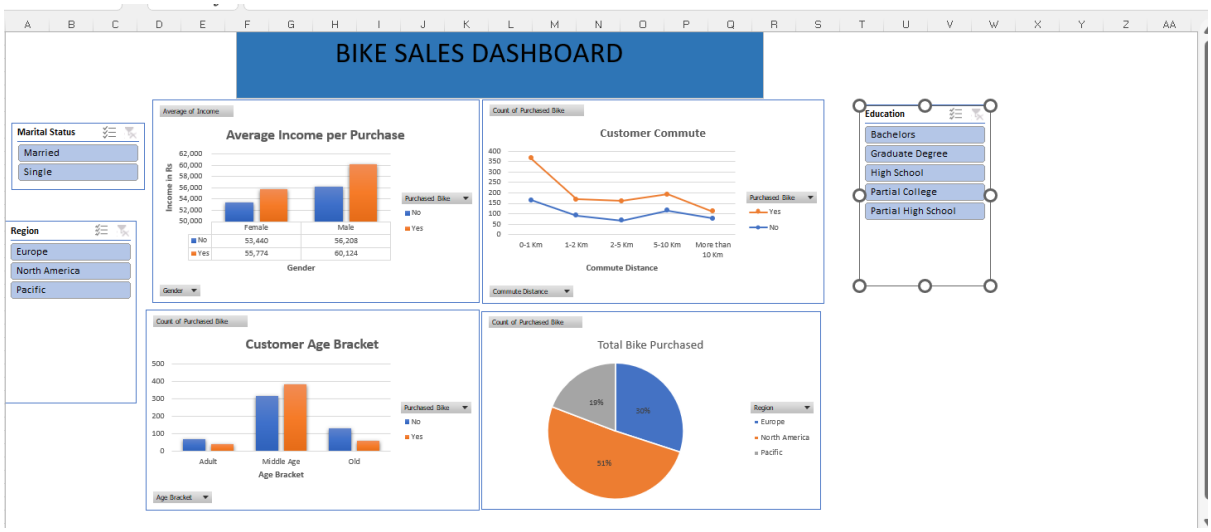


51% bike was purchased from North America Region

Created Dashboard using all 4 charts



Added 3 slicer Region,Education,Marital status and now we can filter the data what we want ,



If we want to see the customer who is in High School ,from North America and single,Filtered the data easily using Slicer to get the desired results.