Exploratory Data Analysis Using Microsoft Excel

Table of Content

Motivation and Aim of Analysis

This project is solely conducted to put my excel skill to test. I am without doubt that during the process I will unveil more abilities and skillsets. The major aim of this analysis is to wrangle and explore this dataset to generate beautiful visualizations, insights and inferences using only Excel.

Information about the Dataset

This Data was obtained from an open-source GitHub account of a youtuber, additionally, it was used during one of my training classes. But the data was not explored in class to my satisfaction simply because everything cannot be taught in class. Which is why I decided to dissect the dataset further and extract more beautiful insights. Also, I was engrossed by this dataset because of its enrichments in both categorical and numerical variables. The raw Data can be downloaded here.

	ID	Marital Status	Gender	Income	Children	Education	Occupation	Home Owner	Cars	Commute Distance	Region	Age	Purchased Bike
2	12496	M	F	$40,000.00	1	Bachelors	Skilled Manual	Yes	0	0-1 Miles	Europe	42	No
3	24107	M	M	$30,000.00	3	Partial College	Clerical	Yes	1	0-1 Miles	Europe	43	No
4	14177	M	M	$80,000.00	5	Partial College	Professional	No	2	2-5 Miles	Europe	60	No
5	24381	S	M	$70,000.00	0	Bachelors	Professional	Yes	1	5-10 Miles	Pacific	41	Yes
6	25597	S	M	$30,000.00	0	Bachelors	Clerical	No	0	0-1 Miles	Europe	36	Yes

This Dataset is from a Bike sale company. It contains information about its customers across three major regions (North America, Europe, and Pacific). It consists of 16 variables(columns) and 1026 observations(rows). Below is the overview of the numerical entries. The maximum, minimum and average values. More information on the categorical variables will be carried out during the exploratory process.

	Income	Children	Cars	Age
Max	$170,000.00	5	4	89
Min	$10,000.00	0	0	25
Average	$56,360.00	2	1	44

Data Wrangling

Primarily, each column is being set to the accurate data type. The numerical columns are set to number, currency, and accounting where necessary. While other categorical variables are set the necessary datatype. The 26 duplicated data are dropped which left 1000 entries in the dataset. The entries in the marital status column are changed Male and Female from M and F which originally sounds confusing with the purchased bike column. New columns are also created to categorize numerical data like number of children, number of cars and age. Below is the comparism between the raw and the cleaned dataset. The edited columns are colored orange while the new created columns are in blue

Before

	ID	Marital Status	Gender	Income	Children	Education	Occupation	Home Owner	Cars	Commute Distance	Region	Age	Purchased Bike
2	12496	M	F	$40,000.00	1	Bachelors	Skilled Manual	Yes	0	0-1 Miles	Europe	42	No
3	24107	M	M	$30,000.00	3	Partial College	Clerical	Yes	1	0-1 Miles	Europe	43	No
4	14177	M	M	$80,000.00	5	Partial College	Professional	No	2	2-5 Miles	Europe	60	No
5	24381	S	M	$70,000.00	0	Bachelors	Professional	Yes	1	5-10 Miles	Pacific	41	Yes
6	25597	S	M	$30,000.00	0	Bachelors	Clerical	No	0	0-1 Miles	Europe	36	Yes

After

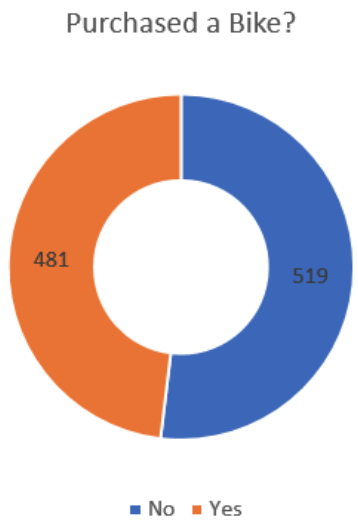
	ID	Marital Status	Gender	Income	Children	Have Children	Education	Occupation	Home Owner	Cars	Own a Car?	Commute Distance	Region	Age	Age Range	Purchased Bike
2	12496	Married	Female	$40,000	1	Yes	Bachelors	Skilled Manual	Yes	0	No	0-1 Miles	Europe	42	Middle Age 31-50	No
3	24107	Married	Male	$30,000	3	Yes	Partial College	Clerical	Yes	1	Yes	0-1 Miles	Europe	43	Middle Age 31-50	No
4	14177	Married	Male	$80,000	5	Yes	Partial College	Professional	No	2	Yes	2-5 Miles	Europe	60	Old 50+	No
5	24381	Single	Male	$70,000	0	No	Bachelors	Professional	Yes	1	Yes	5-10 Miles	Pacific	41	Middle Age 31-50	Yes
6	25597	Single	Male	$30,000	0	No	Bachelors	Clerical	No	0	No	0-1 Miles	Europe	36	Middle Age 31-50	Yes
7	13507	Married	Female	$10,000	2	Yes	Partial College	Manual	Yes	0	No	1-2 Miles	Europe	50	Middle Age 31-50	No
8	27974	Single	Male	$160,000	2	Yes	High School	Management	Yes	4	Yes	0-1 Miles	Pacific	33	Middle Age 31-50	Yes
9	19364	Married	Male	$40,000	1	Yes	Bachelors	Skilled Manual	Yes	0	No	0-1 Miles	Europe	43	Middle Age 31-50	Yes
10	22155	Married	Male	$20,000	2	Yes	Partial High School	Clerical	Yes	2	Yes	5-10 Miles	Pacific	58	Old 50+	No

Exploratory Analysis

In this section, the already wrangled dataset will be explored and visualized to generate insights.

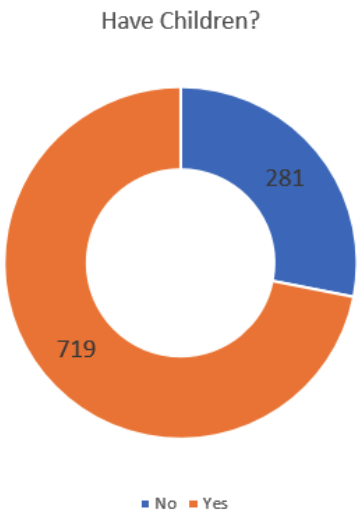
- ## Purchased a Bike?

Only 481 customers bought a bike leaving 519 people without a single bike from this company. Below is the visualization using a pie chart.



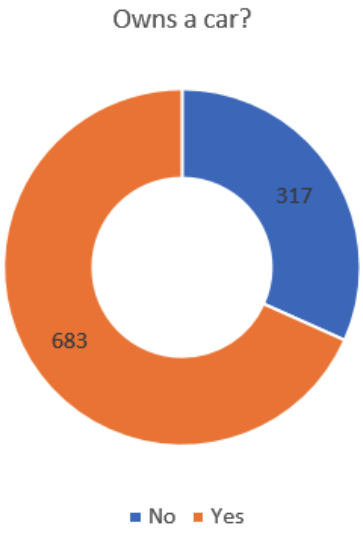
- **Have Children?**

Only 719 customers have at least a child leaving 281 without a child. This is also visualized below using pie chart.



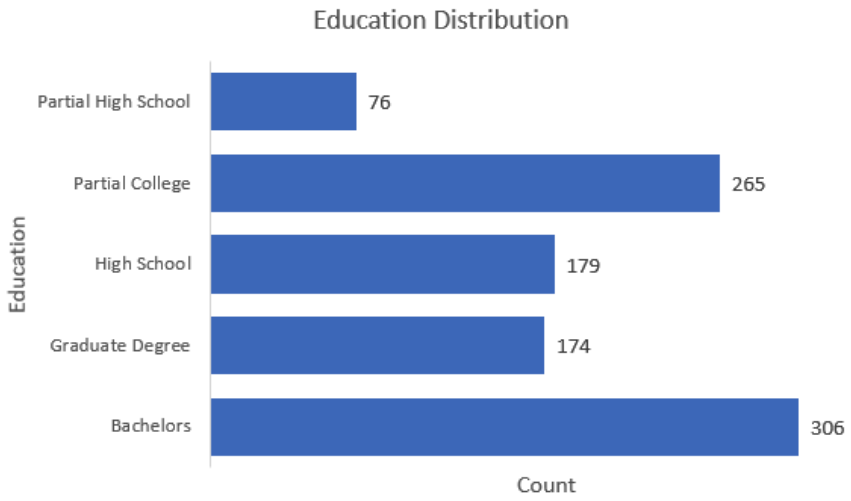
- **Owns a Car?**

Only 683 customers own least a Car while 317 customers have no car.



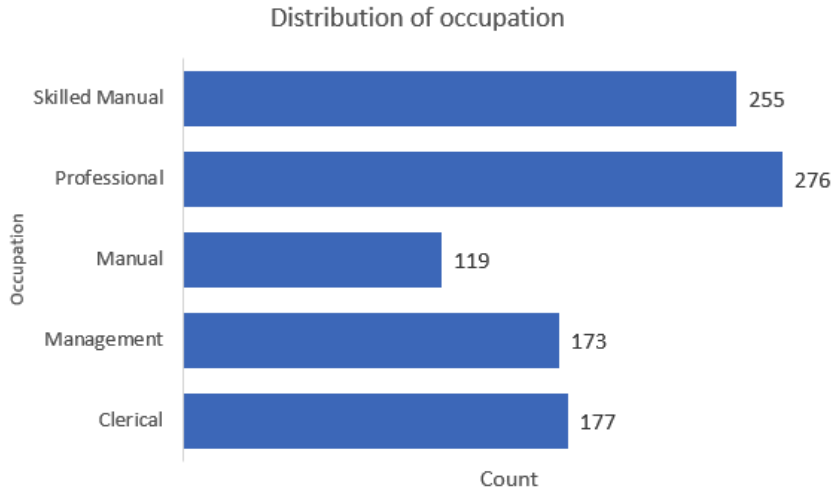
- **Education Distribution:**

The highest number (306) of the entire customers own a bachelor's degree, followed by Partial College (265) and the lowest have Partial High School (76). This is visualized using a bar chart in Excel.



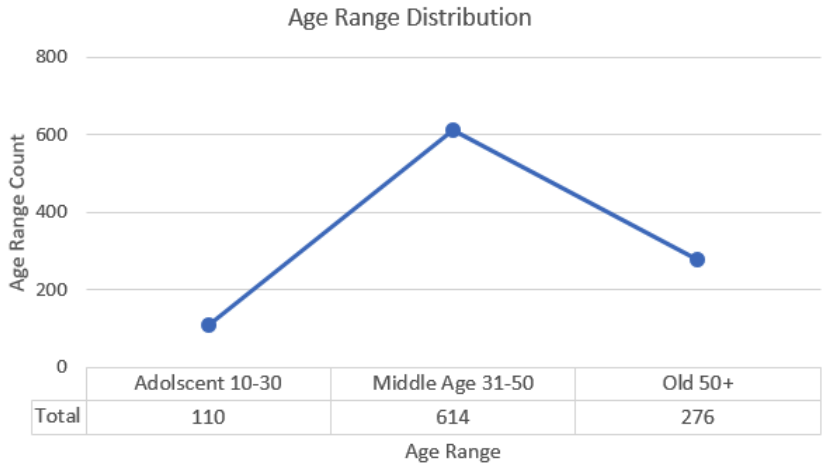
- **Occupation Distribution:**

The total number of 276 customers are Professionals, 255 are Skilled Manual while the least are the customers with Manual Occupations. This is also visualized using a bar chart.



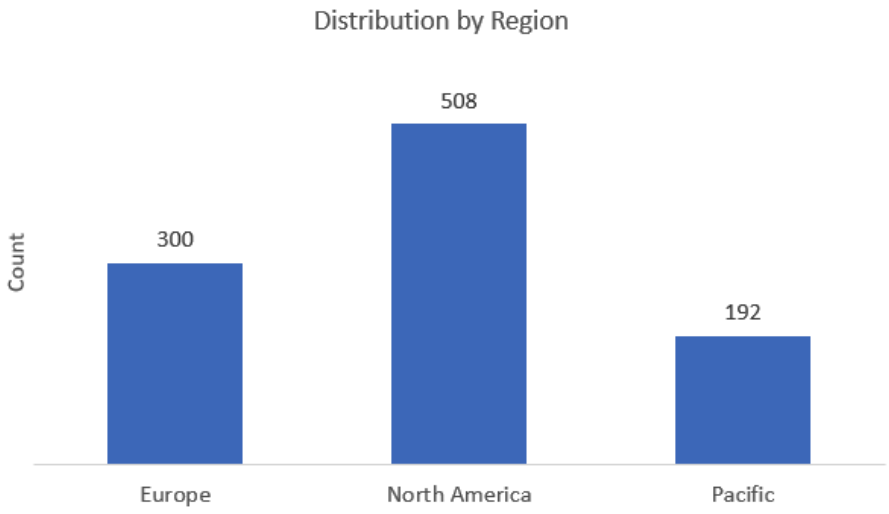
- **Age Range Distribution:**

After categorizing the Age column into ranges, then exploring it gives the following insights. The dataset consists of most customers within the age range of 31-50 which I tagged Middle-Age; total of 614 Customers. Adolescent (10-30) are just 110 and Old (50+) are just 276. This is visualized using a line chart.



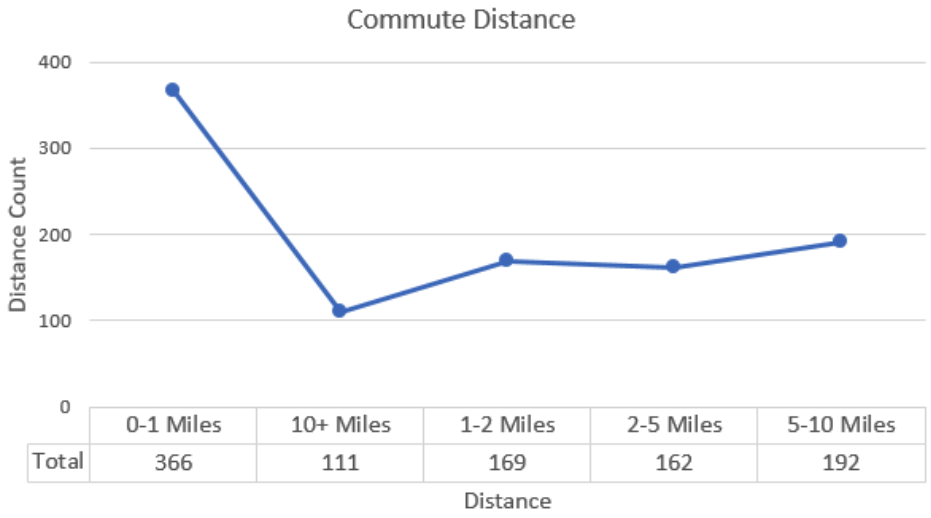
- **Distribution by Region**

Most of the customers are North American (508) followed by European (300) and the least set of customers are from Pacific (192).



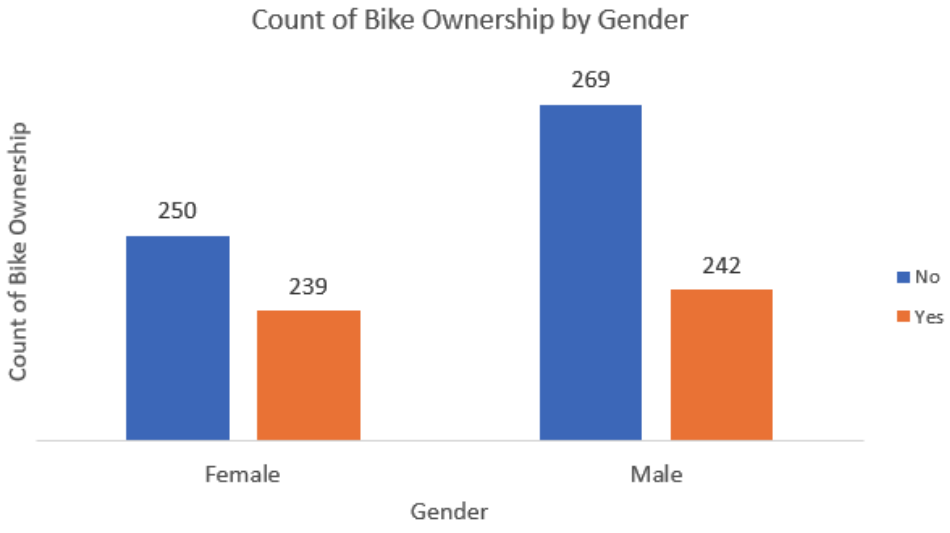
- **Distribution of the Commute Distance**

The largest number of customers (366) commute between 0-1 miles. While the lowest (111) commute 10 miles +.



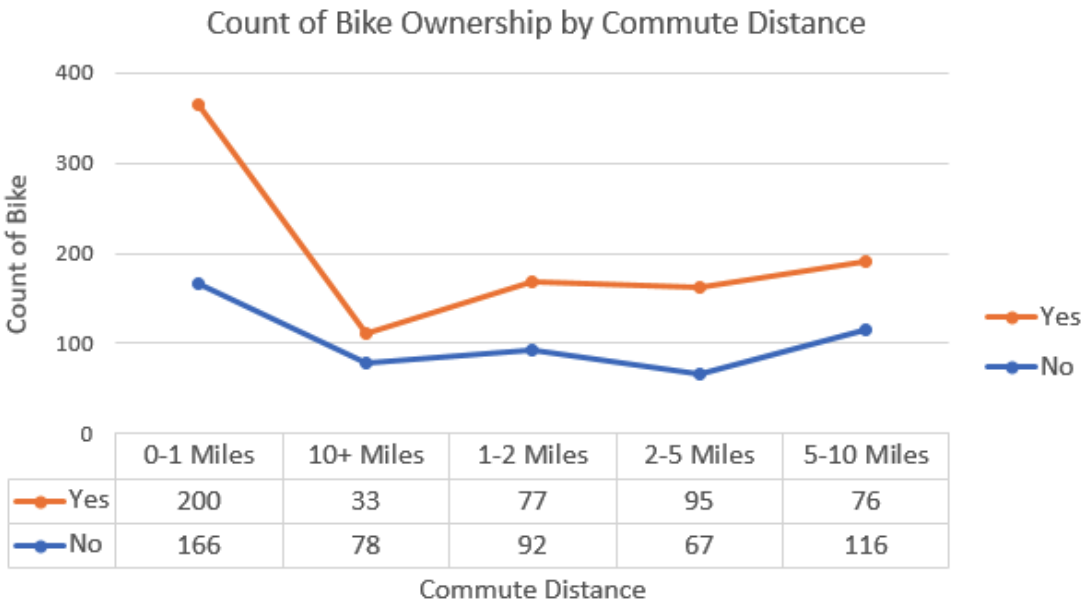
- **Count of Bike Ownership by Gender**

The highest number of customers' gender who buy and do not buy a bike are Male. 242 Male buy a Bike while 269 do not; 239 female customers buy a Bike while 250 do not buy a Bike. Below is the visualization using a column chart.



- **Count of Bike Ownership by Commute Distance:**

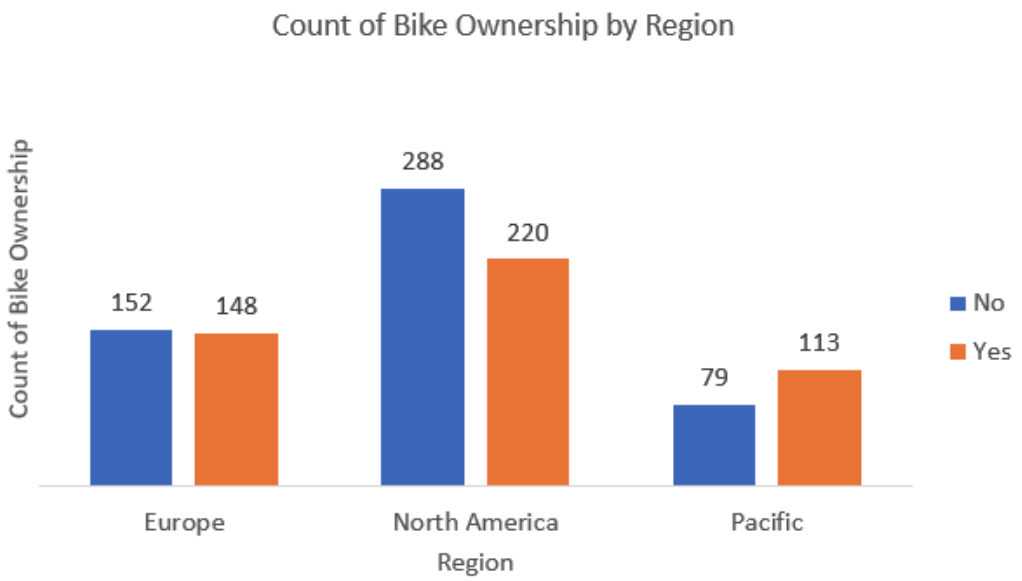
After accessing the dataset, the highest number of customers who purchased a Bike are among those customers commuting between 0-1 Miles which sums up to 200 customers (Purchased a Bike). While the lowest number of customers who purchased a Bike are among those customers commuting 10+ Miles. The same analysis applies to those who do not purchase a Bike. Below is a pictorial graph using a line chart.



	0-1 Miles	10+ Miles	1-2 Miles	2-5 Miles	5-10 Miles
Yes	200	33	77	95	76
No	166	78	92	67	116

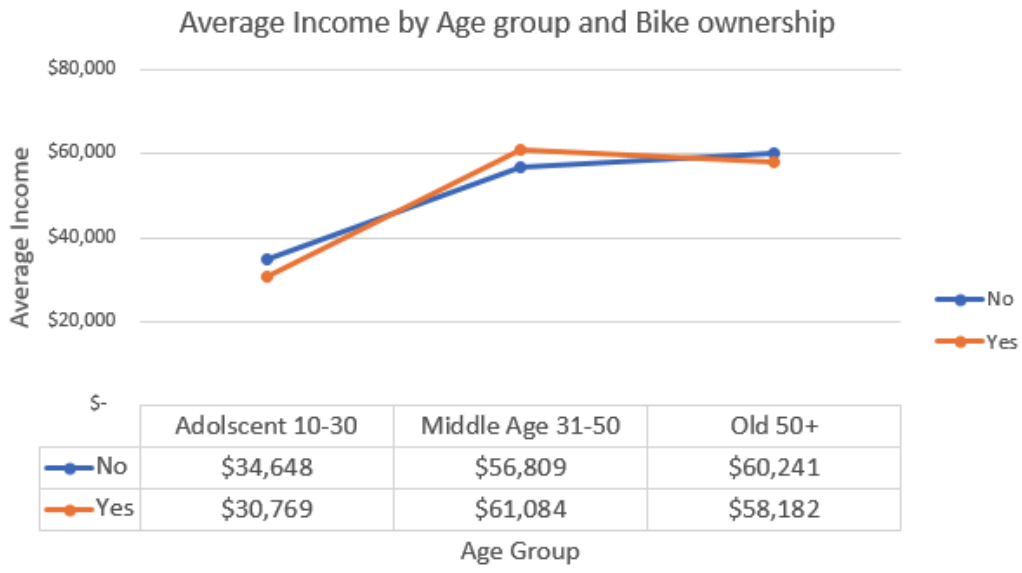
- **Count of Bike Ownership by Region:**

The illustration below shows the distribution of Bike ownership base on region. The highest numbers (220) of customers who purchased a Bike are from North America followed by Europe (148) and lastly Pacific (113). The same goes for customers who do not buy a Bike. 288 North American customers do not buy a Bike, 152 European customers do not buy a Bike and lastly, 79 customers from Pacific do not buy a Bike.



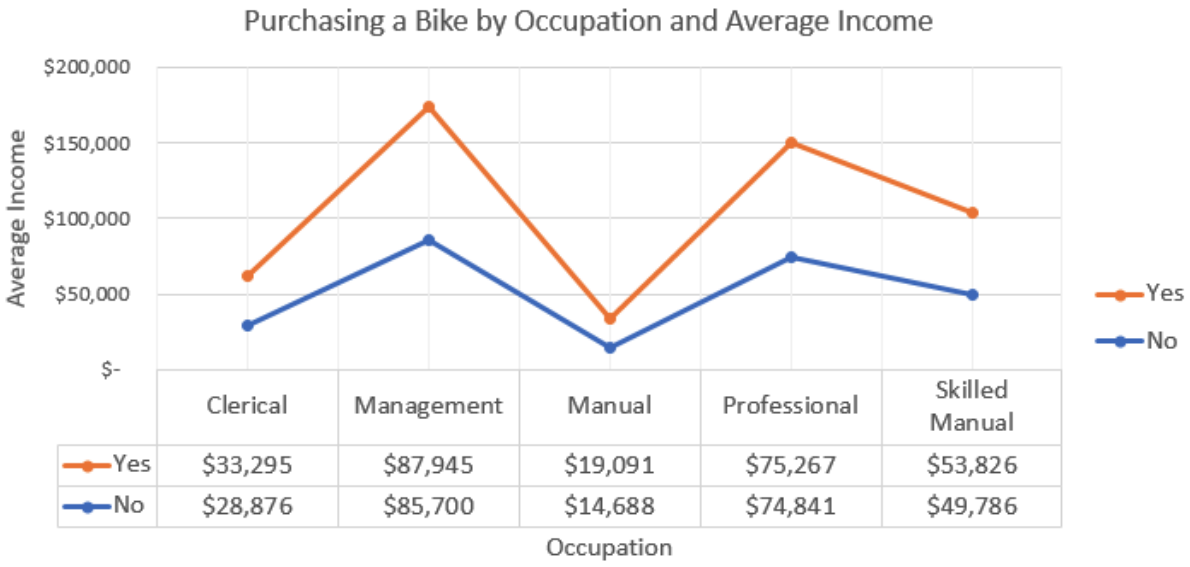
- **Distribution of Average Income by Age Range and Bike Ownership:**

The total average income is 56,360 dollars as illustrated above regardless of whether the customer purchases a Bike or not. Below is the representation of average income of the customers who purchase a bike and do not based on their Age range. Customers between the age of 10-30 have the lowest average income and many of them do not buy a Bike.



- **Distribution of Average Income by Occupation and Bike Ownership:**

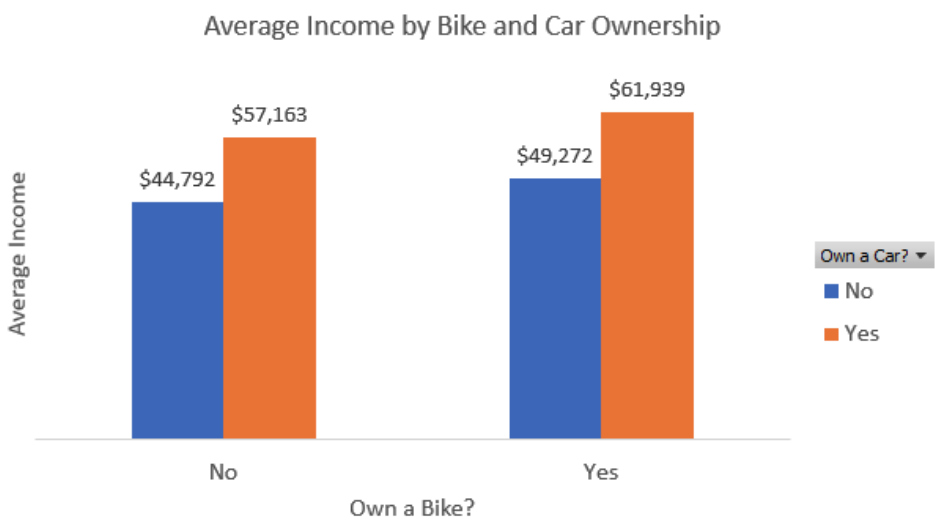
This analysis caught my attention well. The average income across all Occupations here is enough to show the reason behind buying a bike or not. Those customers who buy a Bike have higher average salary than those who do not. Those in Management have a high tendency to own a Bike, of course their income can tell. While those with manual occupation have lowest average salary. The visualization below shows the users with higher average income have more tendency of buying a bike.



	Clerical	Management	Manual	Professional	Skilled Manual
Yes	$33,295	$87,945	$19,091	$75,267	$53,826
No	$28,876	$85,700	$14,688	$74,841	$49,786

- **Distribution of Average Income by Bike and Car Ownership:**

Yes, the doubt still goes on and on. If I have a car, would I need to buy a bike? Regardless, the visualization below shows that most people with a car also have a Bike and a high salary. The Average income of those who own at least one Car and buy a Bike is 61,939 dollars. Which is the highest among the four categories.



Conclusion

As earlier said, the focus of this project is to analyze the Bike sale company's dataset solely using Microsoft Excel. After accessing and wrangling the dataset, 1000 observations were being explored thoroughly. It was derived that there are more customers who do not purchase a Bike. 519 customers do not buy a Bike which is more than 50% of the entire dataset. After looking more inward on what could have caused that, only few Male gender buy a Bike. It is advisable to use a converting advertising strategy when advertising to or engaging the male gender. Also targeting more females as its shown in this dataset that they are the target customers. The number of commute Distance plays a key role when purchasing a Bike. From this analysis, only customers commuting distance between 0-1 Miles buy Bike the most. This set of customers should be targeted. Also, people between the age 30-50 should be targeted for more positive sales conversion. The income of the customers also talks more about the status of their respective pockets. Customers with higher income above 70,000 Dollars on average have greater tendency of purchasing a Bike. What if I own a car, should I still buy a Bike? The dataset shows that the customers with an average income of 61,000 Dollars own at least a car and a Bike. The income of the customer is a significant determining factor when acquiring a Bike.