**Bike Sales Dataset**

**Introduction**

Bikes are human-powered, pedal-driven, single-track vehicle, having two wheels attached to a frame, one behind the other. Riding a bike is fun, it improves your physical and mental wellbeing, is convenient for getting around locally and is environmentally friendly. This project is set out to identify the preference of people between bike and cars.

Data Analysis Phase

ASK 🡪 PREPARE 🡪 PROCESS 🡪 ANALYZE 🡪 SHARE

**ASK**

**Questions:**

1. What is the average income earned by a customer?
2. What factors are significantly related to the number of bikes purchased?
3. Which region has the highest no of bike purchased?
4. Which of the Age Bracket is most interested in purchasing bike?

**PREPARE**

The dataset contains over 1000 values, the dataset was gotten from @Alex the analyst (https://www.youtube.com/watch?v=opJgMj1IUrc&list=PLUaB-1hjhk8H48Pj32z4GZgGWyylqv85f)

It consists of 12 columns which are:

* Customer ID: The unique ID of the customer
* Marital Status: The customers marital status either Married or Single
* Gender: The customers gender
* Income: Customer’s Income ($)
* Children: No of children the customer has (Between 1—5)
* Education: Level of education attained by the customer
* Occupation: Customers occupation
* Home owner: If the customer is a home owner or not
* Cars: Amount of cars owned by the customer (Between 1-4)
* Commute Distance: The distance each customer commutes (miles)
* Region: Customer’s Region
* Age: The Age of the customer
* Purchased Bike: If the customer purchased a bike or not

**PROCESS**

I duplicated the bike dataset and named it working sheet so as to not alter the main dataset.

I checked for duplicated values in the entire dataset by clicking on all columns and 26 duplicated values found leaving 1000 unique values. I changed the M and S to married and single in the marital status column, I changed the income to a datatype of currency from Numbers. I created a new column called “Age Brackets” by grouping the age into >31, <31, and <54 by using the formula

[IF (L2>54,”old”, IF (L2>=31, “middle age”, IF (L2<31, “young”, “invalid”)))

**ANALYZE**

* What is the average income a customer?

The average income of a customer is **$56,360**

* What factors are significantly related to number of bikes purchased?

The age brackets of customers and their marital status

* Which region has the highest purchase of bike?

**North America**

* What is the highest commuted distance?

0-1 miles is the highest commuted distance

**Insights and Recommendations**

* Most single customer’s with a bachelor’s degree purchase more bike than other customers.
* Customers within the Age bracket of middle age tend to purchase more bikes than old or adolescent customers
* The average income of female customers should be increased to allow more females purchase bike
* More enlightment should be done to make people from other regions have interest in purchasing bike.