

# **INTRODUCTION TO DATA MANAGEMENT PROJECT REPORT**

(Project Semester August-December 2022)

## **BIKE BUYERS DATA ANALYSIS**

Submitted by: **Sarvesh Kumar**

Registration No : **12002369**

Programme : **INTRODUCTION TO DATA MANAGEMENT**

Section : **KM006**

Course Code : **INT217**

Under the Guidance of

**Veerpal Kaur : 25909 (Assistant Professor)**

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering)**

**Lovely Professional University, Phagwara**



## **CERTIFICATE**

This is to certify that **Sarvesh Kumar** (student's name) bearing Registration no.**12002369** has completed **INT217** <Course Code> project titled, “**Bike Buyers Data Analysis**” under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor : Veerpal Kaur**

**Designation of the Supervisor : Assistant Professor**

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: **6/11/2002**

## DECLARATION

I, **Sarvesh Kumar** student of **INTRODUCTION TO DATA MANAGEMENT** (Program name) under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: **6/11/2022**

Signature : **Sarvesh Kumar**

Registration No.**12002369**

Name of the student: **Sarvesh Kumar**

## ACKNOWLEDGEMENT

Foremost, it is our privilege to express our sincerest regards, special thanks and gratitude to our teacher **Veerpal Kaur** who gave us the golden opportunity to do this wonderful project on the topic “**Bike Buyers Data Analysis**” which also helped us in doing a lot of research and we came to know about so many new things.

I would like to also thank my own college Lovely Professional University for offering such a course which not only improve my programming skill but also taught me other new technology.

Then I would like to thank my friends who have helped me with their valuable suggestions and guidance for choosing this course.

Finally, I would like to thank my all classmates who have helped me a lot.

## **LIST OF CONTENTS :**

- 1. Introduction**
- 2. Objectives/Scope of the Analysis**
- 3. Source of dataset**
- 4. ETL process**
- 5. Analysis on dataset (for each analysis)**
  - i. Introduction**
  - ii. General Description**
  - iii. Specific Requirements, functions and formulas and prediction models**
  - iv. Analysis results**
  - v. Visualization**
- 6. List of Analysis with results**
- 7. References**
- 8. Bibliography**

## **INTRODUCTION :**

This Project is Bike Buyers Dataset for Exploratory Data Analysis. The dataset which is used in this project is taken from Kaggle and has details of 1000 users from different backgrounds and whether or not they buy a bike. This project is done by Creating pivot tables, Dashboard visualization on Microsoft Excel. Getting to know how many people have purchased bikes mostly as per gender, educational wise, Martial Status.

### **The dataset includes customer -**

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

## **Objectives/Scope of the Analysis:**

1. No. of Purchases by Males and Females.
2. Average Income per purchase.
3. Region wise purchases.
4. Count of Bikes purchased by different age brackets.
5. Sale of bikes in different distance ranges.

## **Source of Dataset:**

This Project is Bike Buyers Dataset for Exploratory Data Analysis. The dataset which is used in this project is taken from Kaggle and has details of 1000 users from different backgrounds and whether or not they buy a bike. The dataset includes customer id, Martial Status, Gender, Income, Children, Education, Occupation, Homeowner, Cars, Commute Distance, Region, Age, Purchased Bike.

**Dataset Link :** [https://www.kaggle.com/datasets/heeraldedhia/bike-buyers?select=bike\\_buyers\\_clean.csv](https://www.kaggle.com/datasets/heeraldedhia/bike-buyers?select=bike_buyers_clean.csv)

## **ETL process**

ETL, which stands for extract, transform and load, is a data integration process that combines data from multiple data sources into a single, consistent data store that is loaded into a [data warehouse](#) or other target system.

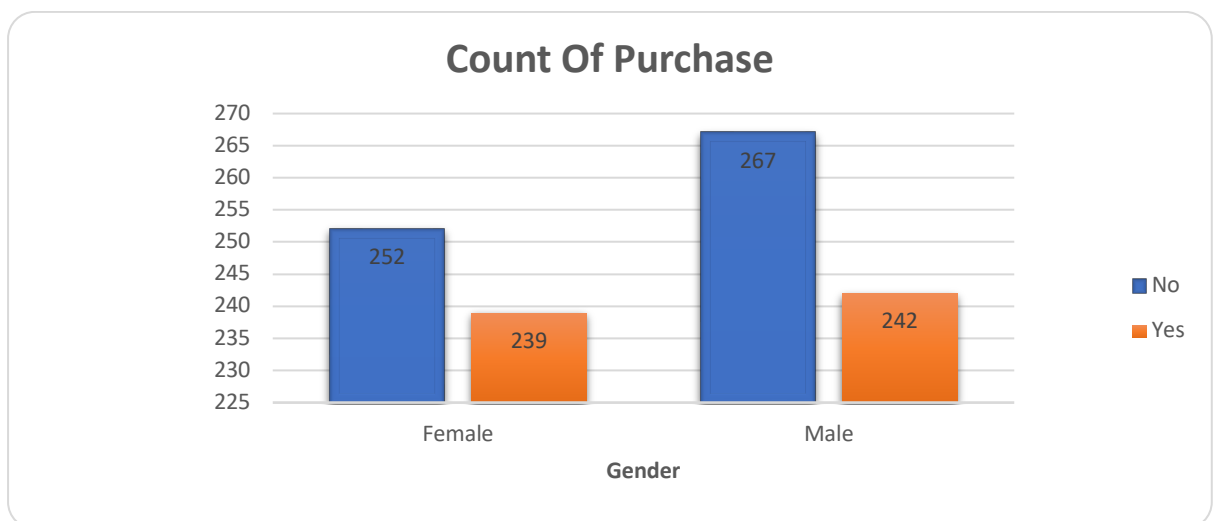
As the databases grew in popularity in the 1970s, ETL was introduced as a process for integrating and loading data for computation and analysis, eventually becoming the primary method to process data for data warehousing projects.

ETL provides the foundation for data analytics and machine learning workflows. Through a series of business rules, ETL cleanses and organizes data in a way which addresses specific business intelligence needs, like monthly reporting, but it can also tackle more advanced analytics, which can improve back-end processes or end user experiences. ETL is often used by an organization to:

- Extract data from legacy systems
- Cleanse the data to improve data quality and establish consistency
- Load data into a target database

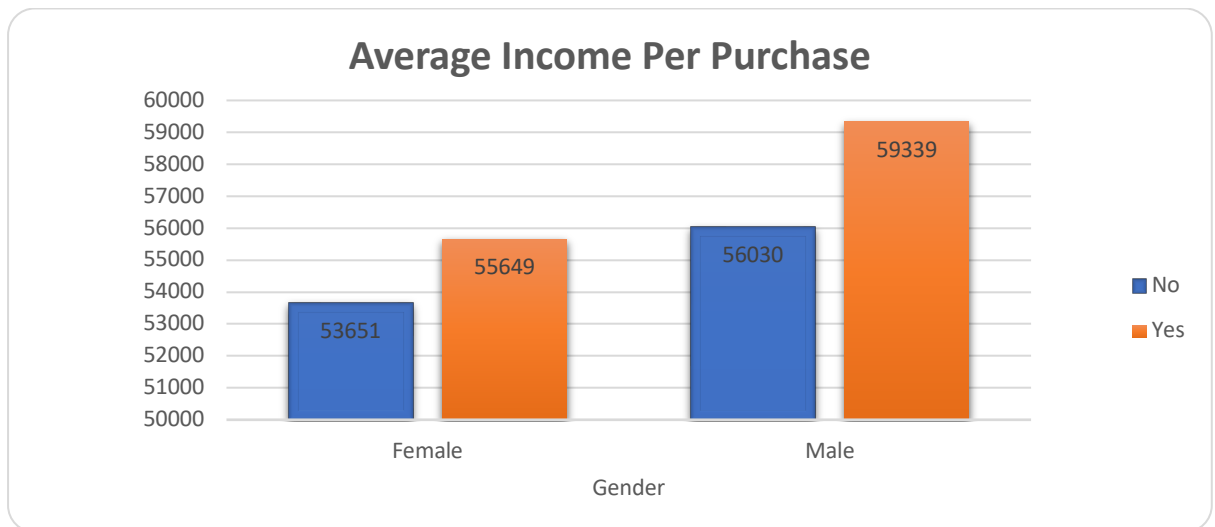
### **Analysis on dataset (for each analysis):**

#### **1. No. of Purchases by Males and Females.**



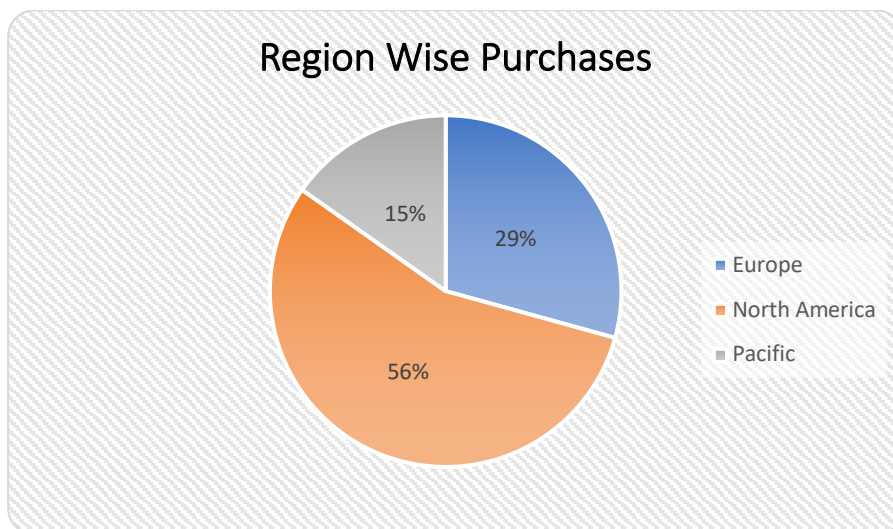
The given graph represents the count of purchases of Bikes by both males and females. After analysing this graph we can say that males purchases the more no. of bikes than female.

## 2. Average Income per purchase.



The given graph represents the Average income per purchase of Bikes by both males and females. After analysing this graph we can say that, Bike is purchased by those who has higher income (around 59,000 or more in males and around 55,000 or more in females). And the average income of male bike owners is higher than females.

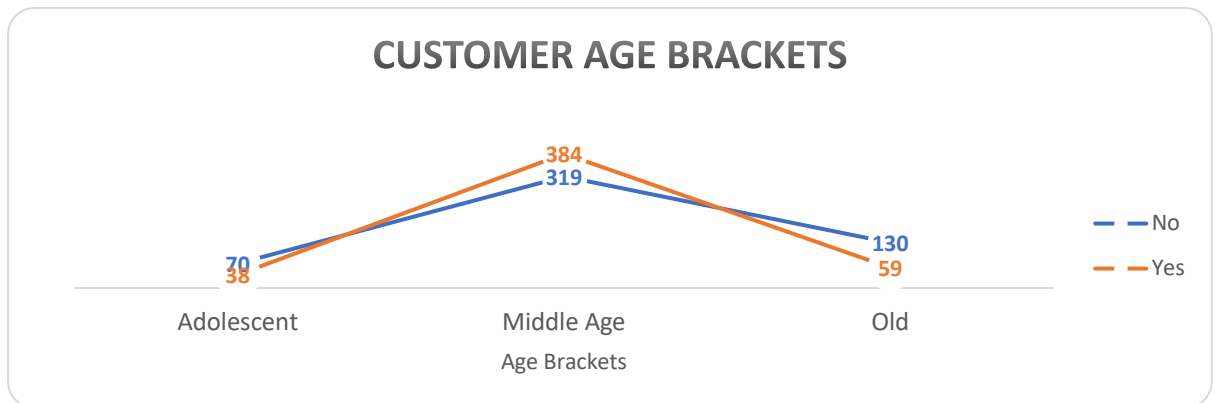
## 3. Region wise purchases.



The given graph represents the Region Wise purchases of Bikes. After analysing this graph we can say that Bike sales in North America is much higher than Europe and Pacific. And Pacific has the least sales.

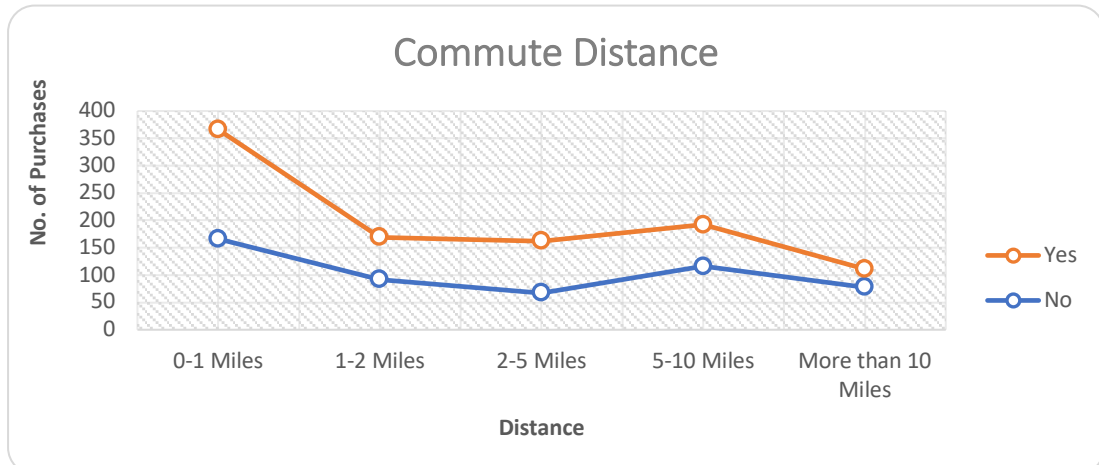


#### 4. Count of Bikes purchased by different age brackets.



The given graph represents the Sale of Bikes in different age brackets. After analysing this graph we can say that Bike sales in Middle Age (31-54) is much higher than Old(>54) and Adolescent(<31). And Adolescent(<31) has the least sales.

#### 5. Sale of bikes in different distance ranges.



The given graph represents the Sale of Bikes in different Commute distance. After analysing this graph we can say that As the distance increases sales decreases. Highest sales of bikes is done in between (0-1) Miles and Least sales is done in (10+) miles .

## **List of Analysis with Results:**

### **1. No. of Purchases by Males and Females.**

After analysis we can say that males purchases the more no. of bikes than female.

### **2. Average Income per purchase.**

After analysis we can say that, Bike is purchased by those who has higher income (around 59,000 or more in males and around 55,000 or more in females). And the average income of male bike owners is higher than females.

### **3. Region wise purchases.**

After analysis we can say that Bike sales in North America is much higher than Europe and Pacific. And Pacific has the least sales.

### **4. Count of Bikes purchased by different age brackets.**

After analysis we can say that Bike sales in Middle Age (31-54) is much higher than Old(>54) and Adolescent(<31). And Adolescent(<31) has the least sales.

### **5. Sale of bikes in different distance ranges.**

After analysis this we can say that As the distance increases sales decreases. Highest sales of bikes is done in between (0-1) Miles and Least sales is done in (10+) miles .

## References :

- [1] <https://books.google.co.in/books?id=CyxjEtnytXUC&lpg=PT4&ots=nSk8Q6j-Yi&dq=data%20management%20using%20excel&lr&pg=PT4#v=onepage&q=data%20management%20using%20excel&f=false>
- [2] [https://www.kaggle.com/datasets/heeraldedhia/bike-buyers?select=bike\\_buyers\\_clean.csv](https://www.kaggle.com/datasets/heeraldedhia/bike-buyers?select=bike_buyers_clean.csv)
- [3] <https://www.youtube.com/watch?v=opJgMj1IUrc>
- [4] <https://www.w3schools.com/EXCEL/index.php>

## Bibliography :

- [https://www.kaggle.com/datasets/heeraldedhia/bike-buyers?select=bike\\_buyers\\_clean.csv](https://www.kaggle.com/datasets/heeraldedhia/bike-buyers?select=bike_buyers_clean.csv)
- Google.
- <https://www.youtube.com/watch?v=opJgMj1IUrc>
- <https://www.w3schools.com/EXCEL/index.php>