# Data Analysis(Part-II)

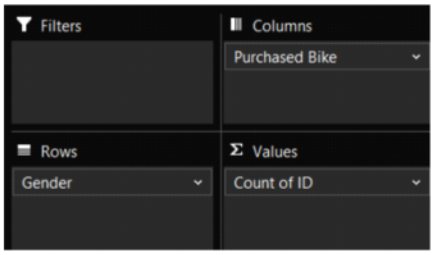
## 🎯 Learning Goals:

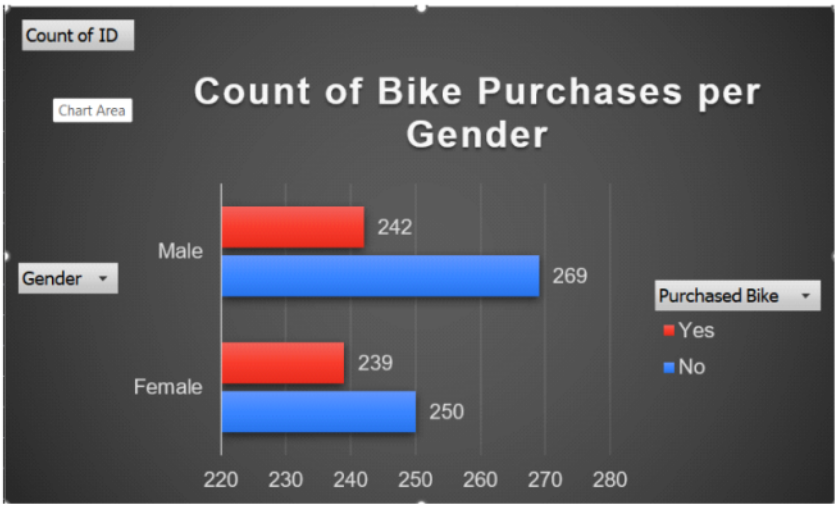
- Learn how to create dashboards using pivot tables and charts
- Understand and apply continuous univariate analysis concepts and Excel tools.
- Understand and apply categorical univariate analysis.
- Use Excel functions like TRIMMEAN(), AVERAGE(), and STDEV() to detect data patterns.
- Use PivotTables, Histograms, and Frequency Tables for visualization.
- Get introduced to the idea of analyzing two variables together (bivariate analysis).

## Challenge 1

### Bar Chart - Count of Bike Purchases per Gender

| Count of ID | Column Labels | | |
|---|---|---|---|
| Row Labels | No | Yes | Grand Total |
| Female | 250 | 239 | 489 |
| Male | 269 | 242 | 511 |
| **Grand Total** | **519** | **481** | **1000** |



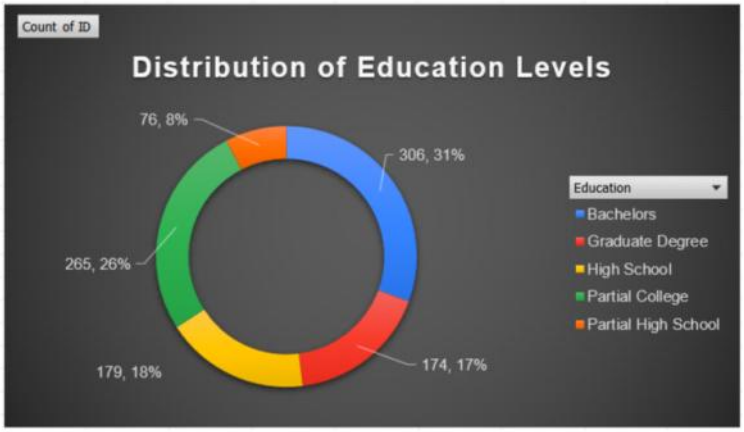



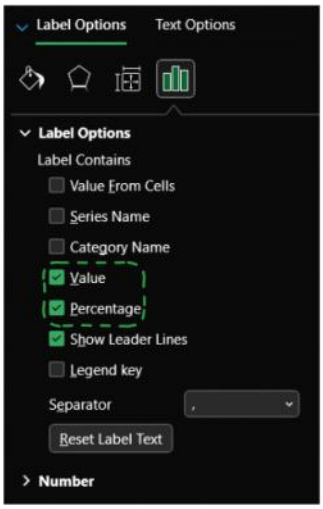
## Challenge 2

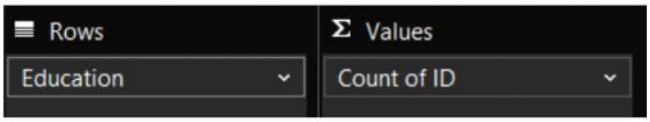
categorical Values

Pie Chart - Distribution of Education Levels

| Row Labels | Count of ID |
|---|---|
| Bachelors | 306 |
| Graduate Degree | 174 |
| High School | 179 |
| Partial College | 265 |
| Partial High School | 76 |
| **Grand Total** | **1000** |

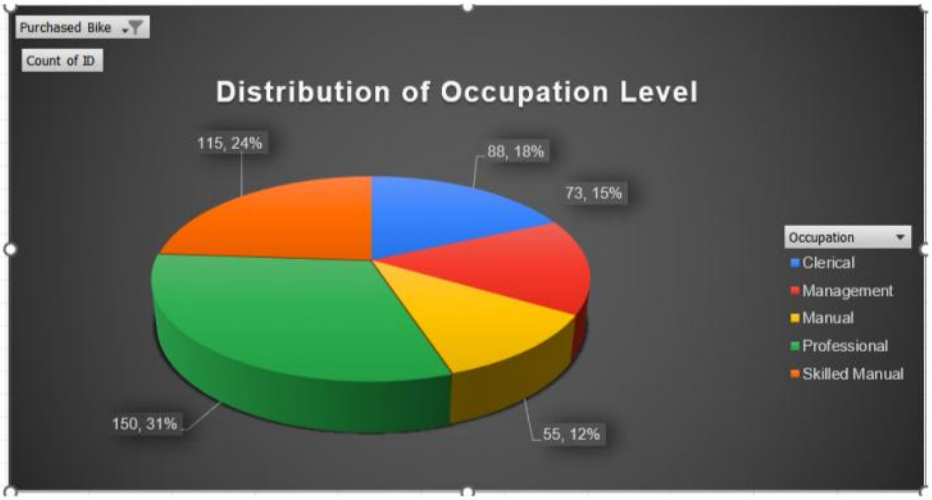


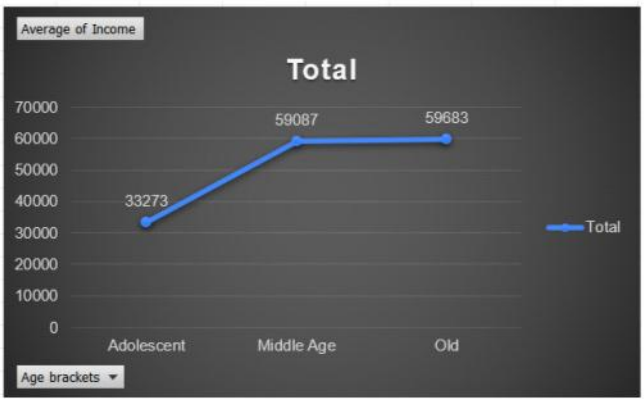

Purchased Bike Yes

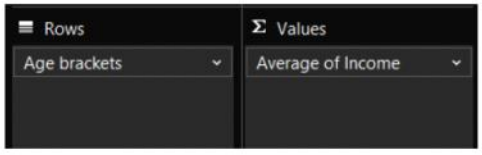
| Row Labels | Count of ID |
|---|---|
| Clerical | 88 |
| Management | 73 |
| Manual | 55 |
| Professional | 150 |
| Skilled Manual | 115 |
| **Grand Total** | **481** |




## Challenge 3

| Row Labels | Average of Income |
|---|---|
| Adolescent | 33273 |
| Middle Age | 59087 |
| Old | 59683 |
| **Grand Total** | **56360** |



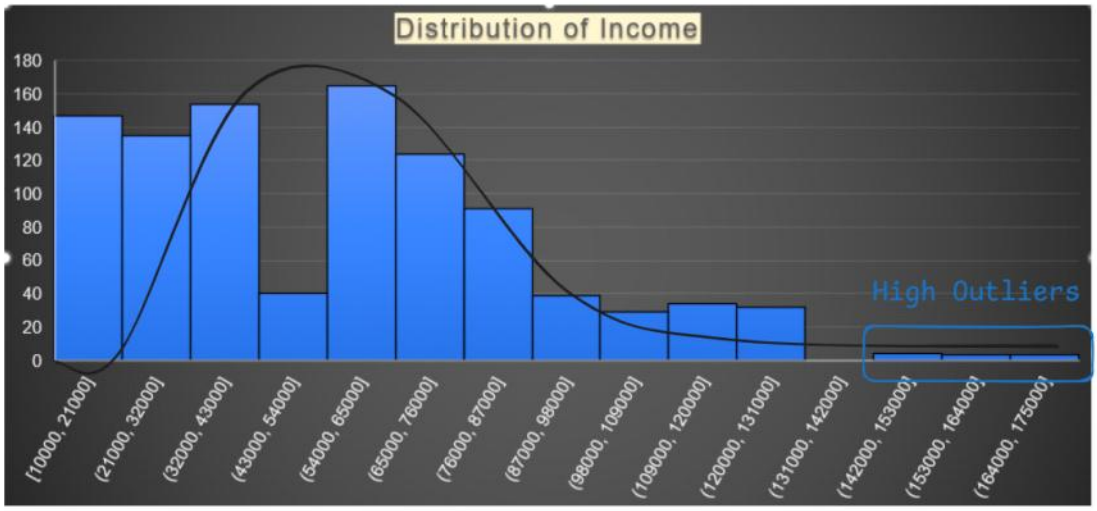



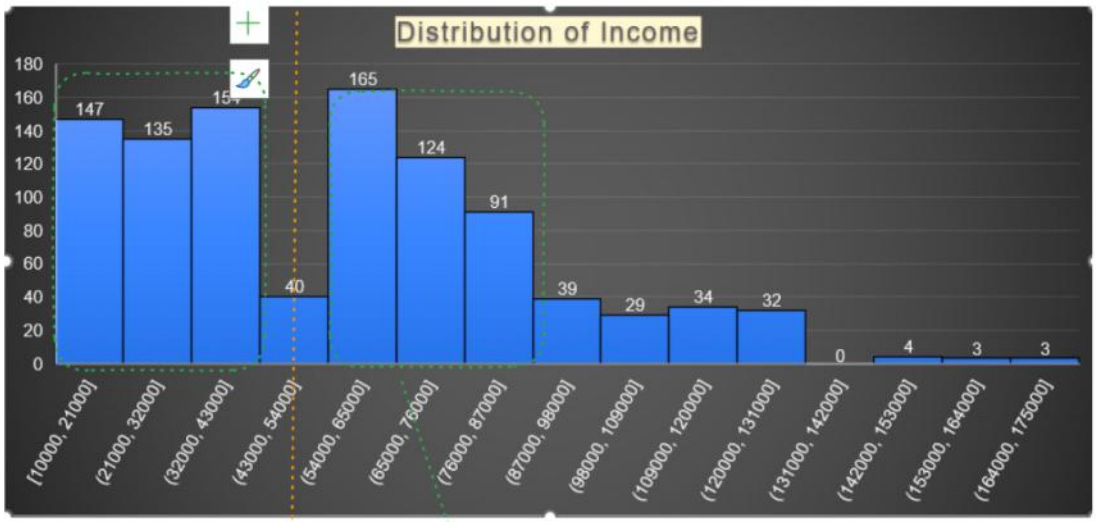
Challenge 4

Histogram - Distribution of Income

- Can't apply histogram on Pivot Table
- Continuous

→ Histogram -> Frequency based distribution -> Normal Distribution Curve -> Bell Curve




10000-43000

54000,87000

A/B Testing

A video

B-Video

## Types of Data :

### 1. Qualitative (Categorical Data)

#### 1.1 Nominal [No natural Orders]

Example - Gender [Male, Female]
- Location [Delhi, Kolkata, Mumbai, Bangalore]
- Product-Category [Clothing, Electronics, Appliances]

#### 1.2 Ordinal [Orders / Ranking]

Example - Customer Satisfaction [Low, Medium, High]
- Membership Tiers [Silver, Gold, Platinum]
- Education [High School, College, Master Degree, PhD]
- Designation[Intern, Associate, Lead, Manager, SeniorManager, Director, V.P]

### 2. Quantitative (Numerical Data)

#### 2.1 Discrete Value [Countable, No Decimal]

Example - Quantity Ordered, No. Of customers, No. of Orders.

#### 2.2 Continuous [Measurable and can have Decimal]

Example - Height, Weight, Age, Income, Profit, Distance, Temp, Product Price, Speed

### Types of Data Analysis

| Types | Description | Examples |
|---|---|---|
| Univariate | Analysis on one variable | - Count Of Gender<br>- Average Income<br>- Sum of Product Price |
| Bivariate | Analysis on two variable | - Bike Purchase by Gender<br>- Population By State<br>- Profit[ProductPrice - ProductCost]<br>- Income Vs Age |
| Multivariate | Analysis on 3 or more variable | - Bike Purchase By Gender + Region + Education |

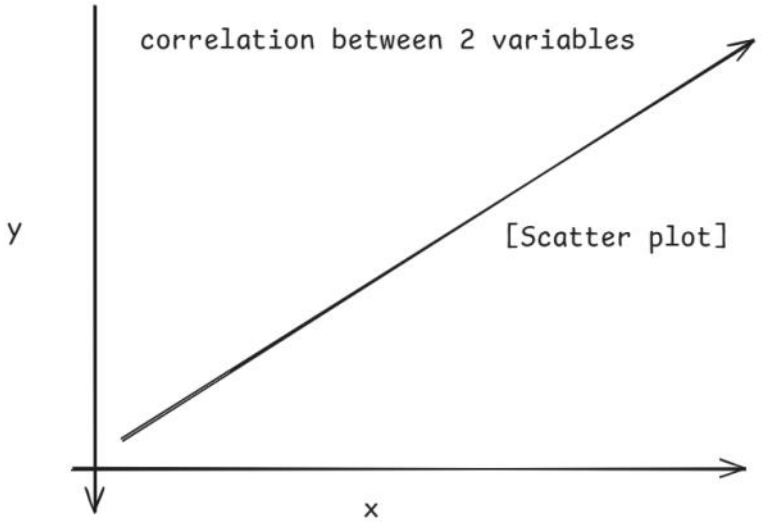
Univariate Analysis → Picking one column at a time/ Analyzing with one variable

| Marital Status | Gender | Income | Children | Education | Occupation | Home Owner | Cars | Commute Distance | Region | Age |
|---|---|---|---|---|---|---|---|---|---|---|

Categories -> Pivot Table
-> Visuals : Bar Chart/Pie Chart
-> Count Frequency, COUNTIF, Percentage/Proportions.

Numerical ->

1. Summary Statistics - Mean/Median/Mode/Std/Variance
2. Visuals -> Histogram , Box Plot




AVERAGE - 53K

Trim Mean VS Average[Mean]

10 records

5K 20K 25K 30K 40K 45K 50K 55K 60K 200K

low outlier High Outlier

TRIMMEAN(range,percentage)

5K-200K , 0.2

TRIMMEAN - 20K - 60 / 8

- 45K

20% of 10 records = 2

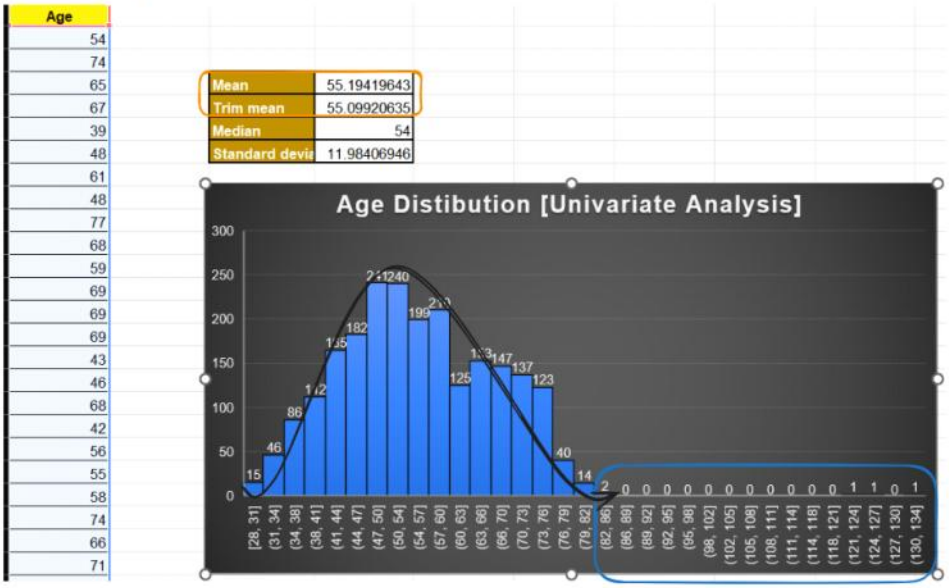
10% remove top 10% remove bottom

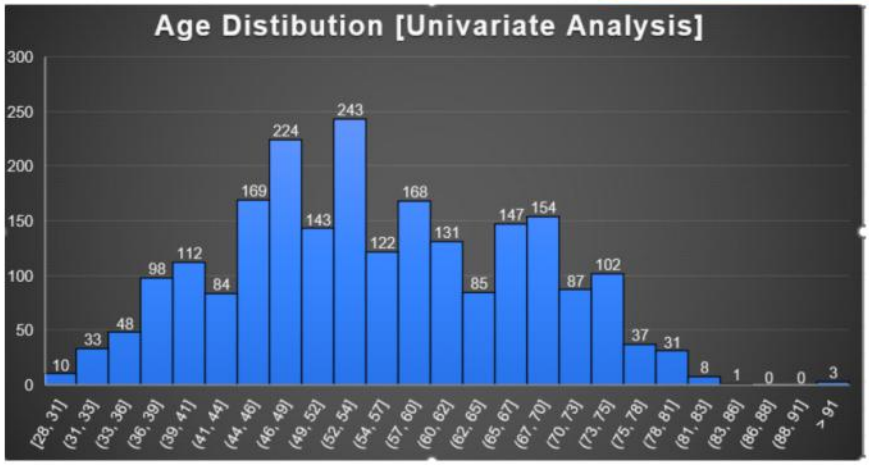
trimMean < Mean ~ Positive Skew [High Outlier]
trimMean > Mean ~ Negative Skew [Low Outlier]
TrimMean ~ Mean -> Symmetric -> [Minimal or no extreme value]

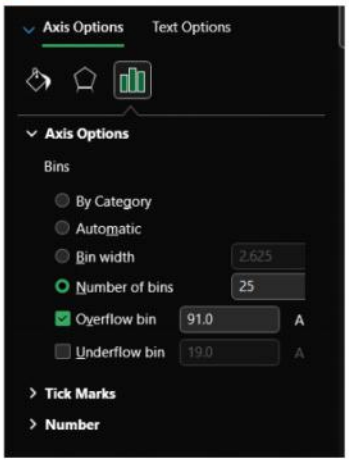
## TrimMean [50K] < Mean [54K] [High Outlier]

| Age |
|---|
| 54 |
| 74 |
| 65 |
| 67 |
| 39 |
| 48 |
| 61 |
| 48 |
| 77 |
| 68 |

| | |
|---|---|
| Mean | 55.19419643 |
| Trim mean | 55.09920635 |
| Median | 54 |
| Standard deviation | 11.98406946 |



| Income |
|---|
| 55158 |
| 52203 |
| 82576 |
| 73803 |
| 7500 |
| 7500 |
| 32632 |
| 75484 |
| 27469 |
| 53230 |
| 64176 |
| 49431 |
| 62972 |
| 62972 |
| 56937 |
| 27683 |
| 22304 |
| 160803 |
| 36778 |
| 43638 |
| 29435 |
| 27203 |
| 68281 |
| 53593 |

| | |
|---|---|
| Mean | 52247.25135 |
| Trim mean | 51779.17313 |
| Median | 51741.5 |
| Standard deviatio | 25037.79717 |