CHAPTER 1

INTRODUCTION

This is the age where vehicle in every community is the most important aspect may it be a bike, scooter, car or anything. This particular aspect of travelling from Place A to Place B is become very important and necessary. The main purpose why travelling though vehicles were to travel from point A to point B. This was the purpose behind the invention of vehicles in the industry and all the places around the world, this particular invention changed the entire world and made it so dependent on vehicles that people use vehicle wherever they go.

The main purpose was to travel from point A to point B, but now as the days went by and the time passed, the need for vehicle changed, for some people it became off-roading, for some it became performance, for some it became racing and so on. This was a massive development as the bikes were also designed as per the needs of the people for people who wanted off roading, bikes were made which completely focused on off roading and which completely gave importance to it, and for some people who wanted more mileage, bikes were designed to give more mileage, this is how the development of bikes became a massive development in the world which changed the lives of people.

After all this developments, when a person wants to take a bike, it has become a challenge as he takes advises from many people with many views like,

“Take a bike with more power”,

“Take a bike with more mileage”,

“Take a bike with more performance”,

These multiple types of suggestions have become a confusing factor for many people as they are not able to decide on what type of bike to take.

This study shows how many people are interested in what kinds of bike and the needs for the features in the bike with different age groups and different genders.

This study shows the similarity within the choices of multiple types of bikes chosen between people of similar age groups and similar genders.

1.1 Motivation

The motivation behind this study arises from the question “ What bike should I buy? ”, “What feature should I focus on? ” and so on. These are the questions which arise in every person’s mind who wants to pick up a bike. The motivations behind this project are:

1. Navigating an Overcrowded Marketplace The contemporary two-wheeler industry has expanded significantly, resulting in a landscape where consumers are presented with an exhaustive list of alternatives. Options now range from high-efficiency commuter vehicles to performance-oriented sports bikes. While this abundance signifies industrial progress, it simultaneously creates a decision-making paralysis for the buyer. The sheer volume of available segments—including cruisers, scooters, and off-roaders—complicates the selection process. A primary driver for this study is the necessity to categorize these options effectively, helping buyers navigate a saturated market without feeling overwhelmed.

2. Overcoming Subjective Influence and Contradictory Guidance A major obstacle in the vehicle purchasing journey is the management of conflicting external advice. Prospective buyers are frequently subjected to a wide spectrum of opinions; family members may advocate for fuel economy and durability, while peers often emphasize speed and visual appeal. This dichotomy—essentially the conflict between practicality (mileage) and aspiration (power)—creates significant confusion. This research is motivated by the need to bypass these subjective, anecdotal opinions and instead focus on objective patterns to determine what truly drives consumer satisfaction.

3. Investigating Demographic Correlations and Stereotypes Purchasing behavior is rarely uniform; it is heavily contingent upon the buyer's profile. There is a critical need to examine how specific variables, namely age and gender, dictate vehicle preference. Conventional wisdom often relies on stereotypes, such as the assumption that younger demographics are solely interested in velocity or that specific genders prefer lightweight vehicles. This study seeks to scientifically validate or debunk these assumptions, analyzing whether these traditional lines are blurring in the modern context or if distinct preference patterns remain intact.

4. Establishing an Empirical Basis for Decision Making Ultimately, the goal of this project is to substitute guesswork with statistical evidence. In a domain often ruled by marketing trends and general assumptions, there is a lack of concrete data regarding local consumer habits. By implementing a structured survey analysis, this project aims to visualize tangible trends in consumer behavior. Transitioning to a data-driven approach allows for a factual understanding of which attributes—be it safety, cost-efficiency, or performance—hold the highest priority for different population segments.

The above mentioned points are the motivations behind this project due to which I have taken up this project and have completed it.

This project is a data analysis project in which the data collected from people all around the world is being cleaned, processed and analysed through Data Visualization.

1.2 Problem Statement

The Dilemma of Choice in a Saturated Market While the evolution of the automobile industry has provided the consumer with an unprecedented variety of options, it has arguably created a complex decision-making environment. In the past, the purchase of a two-wheeler was a straightforward necessity, largely dictated by availability and basic utility. However, the current landscape is characterized by hyper-segmentation. Consumers are no longer just choosing a vehicle; they are choosing between distinct categories such as street-naked, fully faired sports, adventure tourers, cruisers, and commuter scooters. The central problem lies in the "Paradox of Choice"—as the number of options increases, the ease of making a satisfactory decision decreases.

The Conflict of Attributes: Performance vs. Efficiency A significant issue facing prospective buyers is the inherent trade-off between conflicting vehicle attributes. The most prominent of these is the Mileage vs. Power debate. The average consumer often desires a vehicle that excels in all areas—high speed, aggressive styling, maximum fuel efficiency, and low maintenance. However, engineering reality dictates that these features are often inversely proportional. A buyer is frequently forced to compromise, yet they lack the objective data to understand which compromise best suits their specific demographic profile. This leads to confusion, where the buyer vacillates between the practical advice of elders (who prioritize economy) and the aspirational influence of peers (who prioritize performance).

The Information Gap and Subjective Bias Currently, most advice regarding vehicle purchase is anecdotal rather than data-driven. A young student looking for a bike relies on the opinions of friends or family, whose experiences may be outdated or biased towards their own preferences. There is a lack of accessible, localized data that clearly maps out what specific age groups and genders *actually* prefer. For instance, manufacturers and dealerships often operate on historical stereotypes—assuming that female riders prefer lightweight scooters or that older riders are uninterested in aesthetics. Without current data, these assumptions may lead to a mismatch between the consumer’s needs and the products marketed to them.

1.3 Objectives

- To Identify Key Preference Factors: To determine which specific attributes (Mileage, Power, Comfort, Safety, or Aesthetics) are prioritized by consumers when selecting a vehicle.
- To Analyze the Impact of Age on Vehicle Choice: To examine how the age of a consumer influences their preference for specific vehicle categories (e.g., whether younger demographics strictly prefer sports bikes while older demographics prefer commuters).
- To Evaluate Gender-Based Preferences: To study the similarities and differences in vehicle selection between genders, specifically checking for trends in the preference for scooters versus motorcycles.
- To Resolve the "Power vs. Mileage" Conflict: To quantify the trade-off between performance and fuel efficiency, identifying the specific demographic segments that are willing to sacrifice mileage for power, and vice versa.
- To Understand External Influences: To assess the extent to which external factors—such as peer pressure, family advice, and brand reputation—impact the final buying decision.
- To Map Current Market Trends: To visualize the current popularity of different bike segments (such as Off-roading, Racing, and Daily Commute) within the local community.

CHAPTER 2

LITERATURE REVIEW / EXISTING SYSTEM

1. The Evolution of the Two-Wheeler Industry Historically, literature regarding the automotive sector has defined the two-wheeler primarily as a mode of utilitarian transport. Early studies on transport economics emphasized the role of the motorcycle as a cost-effective solution for mobility, particularly in developing economies. However, modern research indicates a paradigm shift. According to recent automotive market analyses, the vehicle has evolved from a basic necessity (Place A to Place B) into a lifestyle product. The emergence of sub-segments—such as adventure touring, café racers, and naked sports bikes—suggests that manufacturers are no longer selling just "transportation," but rather an "experience" or "identity."

2. Factors Influencing Consumer Purchase Decisions Standard marketing theories (such as Kotler’s Consumer Behavior Model) suggest that buying decisions are influenced by Cultural, Social, Personal, and Psychological factors.

- Economic Factors: Numerous studies emphasize that Fuel Efficiency (Mileage) remains the dominant deciding factor for the average middle-class consumer. The rising cost of fuel has solidified mileage as a primary metric for evaluation.
- Performance Factors: Conversely, automotive journals highlight a growing segment of "enthusiast buyers" who prioritize Power (BHP and Torque) over economy. This literature points to a conflict in the market where consumers often want the "best of both worlds," leading to confusion during the purchase process.

3. The Role of Demographics: Age and Risk Appetite Psychological studies on risk-taking behavior suggest a strong correlation between age and vehicle preference.

- Youth: Literature on youth psychology suggests that younger demographics (18–25 years) are more prone to risk-taking and status-seeking behavior. Consequently, they are more likely to prefer high-performance, aesthetically aggressive motorcycles.
- Older Demographics: In contrast, studies focus on the shift towards "functionalism" as consumers age. Older demographics (40+ years) tend to prioritize comfort, ergonomics, and reliability over top speed. This project aims to validate if these established theories still hold true in the current local market.

4. Gender Dynamics in Vehicle Ownership Traditionally, the two-wheeler market was heavily segmented by gender, with "Geared Motorcycles" marketed to men and "Gearless Scooters" marketed to women. However, contemporary gender studies in marketing suggest a blurring of these lines. The rise of female riders in the heavy-bike segment and male preference for high-performance maxi-scooters indicates that gender stereotypes may be outdated. Existing literature calls for newer data to understand these changing dynamics.

CHAPTER 3

DATASET OVERVIEW

The dataset used in this study used consists of the survey data collected from people around the world. The people who attended this survey belong to different age groups and different genders who have submitted their preferences based on their choices and their preferences on taking bikes.

The survey was conducted through google forms.

This is the link to the google forms: https://forms.gle/cEUF5guYzA1ARsjE7

The Variables used for this project are:

1. Demographic & Status Variables

- Age Group: Categorizes the respondent into specific age cohorts to analyze generational trends.
- Gender: Used to identify distinct preference patterns between male and female respondents.
- Two-Wheeler Ownership: A binary or categorical variable confirming if the respondent currently owns a vehicle, establishing their experience level.

2. Preference & Behavioral Variables

- Suitability Perception: A column recording which bike type the respondent believes best fits their own age group.
- Reasoning: A qualitative column capturing the "Why"—the underlying reason for their specific choice (e.g., comfort, speed).
- Feature Priority: Identifies the single most important attribute (e.g., Mileage, Looks, Power) influencing the final purchase decision.
- Affordability: Represents the price range the respondent deems appropriate for their specific age bracket.

3. Cross-Generational Perception Variables A significant portion of the dataset includes perception-based columns where respondents recommended the "Best Bike Type" for different age segments:

- Below 20 Years (Students)
- 21–30 Years (Working Professionals)
- 31–40 Years (Family-Oriented)
- 41–50 Years (Mid-Age Riders)
- 51 Years & Above (Senior Riders) *These columns allow for an analysis of societal stereotypes regarding age and vehicle usage.*

4. Future Trends

- Electric Vehicle Opinion: A variable gauging the public sentiment towards Electric Bikes as a viable option for the future.

Variable / Column Name	**Description**	**Purpose in Analysis**
Age Group	The age range of the respondent.	Independent variable to correlate with preferences.
Gender	The gender identity of the respondent.	To check for gender-based purchasing trends.
Current Ownership	Does the user currently own a two-wheeler?	Filters experienced riders vs. non-riders.

Variable / Column Name	**Description**	**Purpose in Analysis**
Self-Suitability	Which bike type suits the respondent's age?	Measures self-perception.
Age-Specific Recommendations	Five separate columns asking for the best bike for ages <20, 21-30, 31-40, 41-50, and 50+.	To map the lifecycle of vehicle preference across a lifespan.
Key Feature	The most important feature when buying (Mileage, Power, etc.).	Identifies the primary "Purchase Driver."
Price Range	The affordable budget for the respondent.	correlates economic status with bike choice.
EV Perception	Opinion on Electric Bikes.	Gauges market readiness for sustainable transport.

To achieve the research objectives, the dataset was structured around 13 core variables. The first section recorded demographic data (Age, Gender) and Ownership Status to establish a baseline profile. The core of the dataset focused on psychographic profiling, asking respondents to identify the Key Features (such as mileage or performance) that drive their decisions, as well as their Budget Constraints. Uniquely, the dataset also captured societal perceptions by asking respondents to map specific bike types to five distinct age brackets (from "Below 20" to "51+"), allowing for a comprehensive analysis of how vehicle needs change over a human lifecycle. Finally, the dataset included a forward-looking variable regarding the acceptance of Electric Vehicles.

3.1 Data Collection Method

The key steps included are:

1. Generated questions for this survey
2. Created a google form for responders to respond to this survey
3. Circulated this survey through social media to multiple people around the world
4. Cleaning done using Python and verified in Power Query
5. Final cleaned dataset used for visualization.

CHAPTER 4

TOOLS AND TECHNOLOGIES USED

The successful execution of this mini-project required a combination of software tools, analytical libraries, data-processing frameworks, and a structured technical environment. These tools enabled smooth data cleaning, transformation, modeling, visualization, analysis, and interpretation. Each tool was selected for its ability to handle large datasets, produce accurate insights, and support both statistical and visual analysis.

4.1 Software Used

1. Google Forms:

To facilitate efficient and widespread data collection, this study utilized **Google Forms**, a cloud-based survey administration software offered by Google. This tool was selected for its versatility in designing structured questionnaires and its ability to reach a diverse demographic through digital dissemination.

The survey questions (as detailed in Chapter 3) were digitized into a user-friendly interface, allowing respondents to submit their answers remotely via smartphones or computers. This digital approach ensured **data integrity** by preventing manual entry errors, as responses were automatically aggregated into a centralized database (Google Sheets) in real-time. The platform also enabled the use of mandatory fields, ensuring that critical data points—such as age group and key preference features—were not left blank by respondents.

2. Python:

Tools Used: Python served as the primary engine for this project, handling the entire data pipeline from raw input to final visualization. Specifically, it was used to:

- Import the raw CSV data exported from Google Forms.
- Execute logic to categorize respondents (e.g., grouping ages).
- Run the web server required to display the dashboard.

3. Pandas (Library):

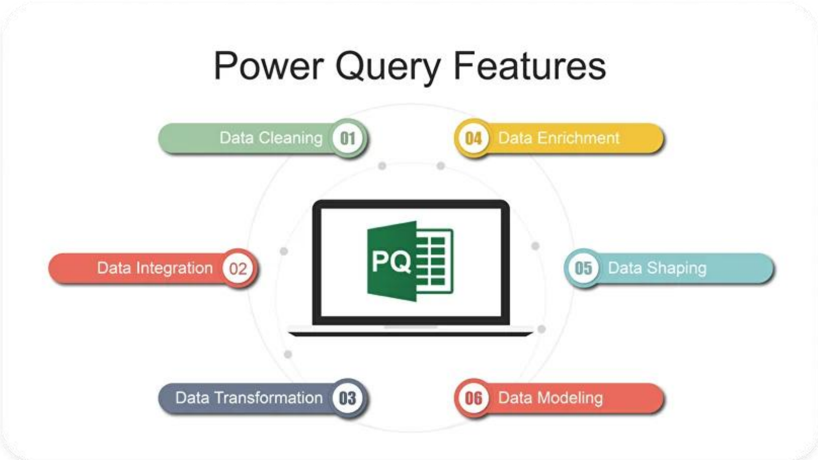
Tools Used: Pandas Pandas was used as the "Excel" of the coding environment. It allowed the raw survey data to be loaded into a DataFrame, where specific operations were performed:

- Cleaning: Removing empty rows or correcting spelling errors in the "Reasoning" column.
- Filtering: Creating smaller sub-datasets (e.g., a specific table just for "Student" responses) for targeted analysis.
- Grouping: Calculating the average budget per age group.

4. Power Query:

Tools Used: Power Query Power Query was used for the initial "rough cleaning" of the data before coding began. It helped in:

- Splitting Columns: Separating complex survey answers where respondents selected multiple options.
- Data Type Assignment: Ensuring price columns were formatted as currency and age columns as numbers.
- Error Removal: Automatically filtering out incomplete survey responses before saving the final CSV file.



5. Plotly Express:

Tools Used: Plotly Express Plotly Express was the tool responsible for converting rows of numbers into visual insights. It was used to generate:

- Interactive Bar Charts: Showing the comparison of "Power vs. Mileage" across different age groups.
- Pie Charts: Visualizing the gender distribution of the respondents.
- Box Plots: Displaying the range of affordable budgets for students versus working professionals.

6. **Flask (Web Framework):**

   Tools Used: Flask Flask acted as the "backbone" of the project application. While not visible to the user, it worked in the background to:

- Launch the local web server so the dashboard could be viewed in a browser.
- Handle the routing (URLs) to ensure the correct charts loaded when the page was refreshed.
- Serve the CSS and HTML files required for the design.

7. **Dash (Dashboard Library):**

   Tools Used: Dash Library Dash was the tool used to build the actual "Website" or "Dashboard" where the results are displayed. It combined everything into one view:

- Layout: It arranged the text, graphs, and images on the screen (e.g., putting the Pie Chart next to the Bar Chart).
- Interactivity: It enabled the dropdown menus, allowing users to filter the charts (e.g., selecting "Female" to see only female preferences).

CHAPTER 5

METHODOLOGY / DATA ANALYSIS PROCESS

5.1 Data Collection

The data used in this project was collected from a survey conducted using Google Forms and the forms were passed over through social media such as Instagram , Whats App, Facebook, Linkedin and Reddit.

5.2 Key Details of Data Collection & Utilization

1. Data Source

- Method: Primary Data Collection
- Platform: Google Forms
- Type: Structured Survey Data
- File Format: .CSV (Comma Separated Values) export
- Access Method: Direct extraction from Google Forms response sheet

2. Dataset Characteristics

The dataset comprises primary responses collected from the local community, focusing on 13 key variables including:

- Demographics: Age Group, Gender, Employment Status.
- Preferences: Key decision factors (Power vs. Mileage vs. Comfort).
- Behavioral: Current ownership status and budget constraints.
- Perception: Age-specific bike recommendations (e.g., "Best bike for students") and Electric Vehicle (EV) adoption willingness.

3. Data Authenticity & Reliability

Since the dataset was collected directly by the researcher (Primary Data):

- Relevance: The data is highly specific to the local market and current consumer trends, unlike generic global datasets.
- Validation: Responses were manually screened to remove duplicates or incomplete entries during the cleaning phase.

- Currentness: The data reflects real-time consumer sentiment for the current year, ensuring the insights are up-to-date.

4. Reason for Choosing Primary Data (Google Forms)

- Tailored Objectives: Public datasets (like Kaggle) did not contain the specific "Age vs. Feature Preference" variables required for this study.
- Customization: Allowed for specific questions regarding "Perceived Suitability" across different age groups.
- Structured Output: The form enforced data types (e.g., restricted age ranges), reducing the need for extensive error correction.
- Ease of Integration: The resulting CSV format is natively compatible with Python Pandas for immediate analysis.

5. Data Processing & Visualization Flow

The collected data was processed through a Python-based pipeline to create the interactive dashboard:

- Excel / Power Query: Used for initial formatting and standardizing column headers.
- Python (Pandas): Used for advanced data cleaning, filtering demographic groups, and statistical aggregation.
- Plotly Express: Used to generate interactive charts (Sunbursts, Bar Charts) that allow hovering and filtering.
- Flask & Dash: The final destination where data was deployed as a live, interactive web application, allowing users to dynamically explore the survey results.

5.3 Data Cleaning

Data cleaning was done using Python Library Pandas the code used to clean the data is below:

```
drop_cols = ['Enter your name', 'Enter your email id']
df = df.drop(columns=[c for c in drop_cols if c in df.columns], errors='ignore')
```

CHAPTER 6

SYSTEM DESIGN/ ARCHITECTURE

The system architecture for this data analysis project defines the end-to-end flow of data from collection to final dashboard creation and analytical interpretation. Although the project is primarily analytical rather than software-oriented, a structured architecture ensures that the workflow remains systematic, replicable, and efficient.

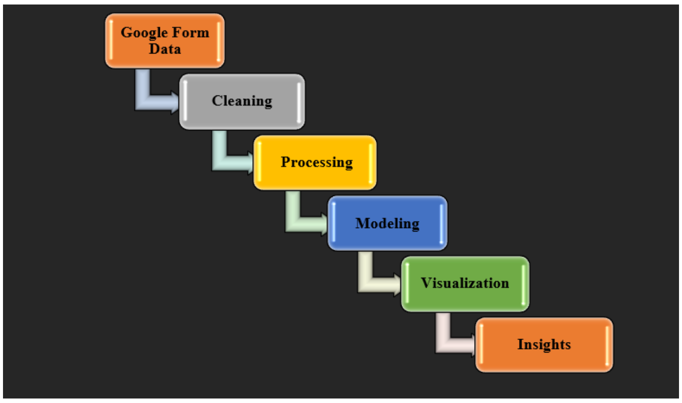
6.1 Overall Architecture Overview

The system architecture follows a **linear analytical pipeline**, starting from dataset acquisition (Google Forms) and ending in dashboard visualization (Flask Dashboard) and analytical validation (Python).

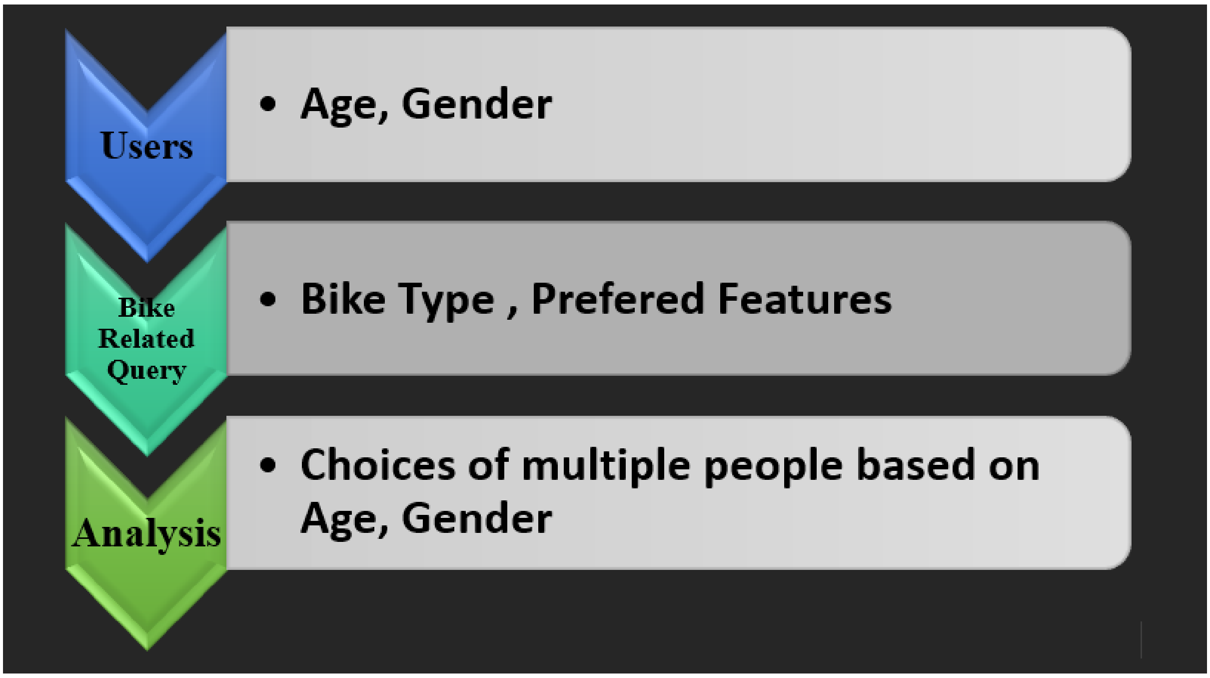
Architecture Stages:

1. **Data Acquisition** – Survey from Google Forms
2. **Data Cleaning** – Remove duplicates, standardize formats
3. **Data Loading** – Import the csv from forms into Python
4. **Visualization** – Dashboard, charts
5. **Analysis & Interpretation** – Comparison of Python & Power BI results
6. **Reporting** – Document preparation with insights, charts, and conclusions

6.2 Workflow Diagram



6.3 ER Diagrams



CHAPTER 7

IMPLEMENTATION

7.1 Implementation Overview

This project moves beyond static analysis by implementing a dynamic web-based dashboard. The implementation does not rely on drag-and-drop BI tools but is built using a code-first approach involving:

- Programmatic Data Cleaning: Using Python scripts to handle missing values and standardize text inputs.
- Data Frame Transformation: Restructuring survey data for analysis using Pandas.
- Interactive Visualization: generating dynamic charts that respond to user inputs.
- Web Deployment: Hosting the analysis on a local Flask server for interactive exploration.

7.2 Code Modules Explanation

1. Data Loading & Preprocessing (Pandas) Instead of a database schema, this project uses a DataFrame structure.

- **Data Ingestion:** Reading the raw CSV file using pd.read_csv().
- **Data Cleaning:** Handling null values and renaming complex column headers from Google Forms into simplified variables (e.g., changing "What is your age?" to Age_Group).
- **Categorization:** Grouping qualitative responses (like "Reason for buying") into standardized categories for better aggregation.

CHAPTER 8

CODE USED

PYTHON – FLASK CODE USED

```
# Bike Preference Survey Python Dashboard
import pandas as pd
import plotly.express as px
import dash
from dash import dcc, html
from dash.dependencies import Input, Output

# Load and clean dataset
file_path = "Bike_preference_survey_updated.csv"
df = pd.read_csv(file_path)

# Drop unneeded columns (you already removed Timestamp)
drop_cols = ['Enter your name', 'Enter your email id']
df = df.drop(columns=[c for c in drop_cols if c in df.columns], errors='ignore')

# Define important columns manually (for reliability)
age_col = 'What  Age Group do you fall in'
gender_col = 'Kindly mention your gender'
ownership_col = 'Do you currently own a two-wheeler?'
pref_col = 'According to your age group, which bike type is most suitable?  '
brand_col = 'When choosing a bike, which feature is most important to you?  '
usage_col = 'Do you currently own a two-wheeler?'  # reuse as "purpose/usage"

# Initialize Dash App
app = dash.Dash(__name__, title="Bike Preference Survey Dashboard")
```

```
# Dashboard Layout
app.layout = html.Div([
    html.H1("🏍️ Bike Preference Survey Dashboard",
            style={'textAlign': 'center', 'color': '#007BFF'}),

    # Filters
    html.Div([
        html.Div([
            html.Label("Filter by Gender:"),
            dcc.Dropdown(
                options=[{'label': g, 'value': g} for g in
sorted(df[gender_col].dropna().unique())],
                id='gender_filter',
                placeholder='Select Gender',
                multi=True
            ),
        ], style={'width': '40%', 'background': 'white', 'boxShadow': '0 3px 8px
rgba(0,0,0,0.1)', 'display': 'inline-block', 'padding': '10px'}),

        html.Div([
            html.Label("Filter by Age Group:"),
            dcc.Dropdown(
                options=[{'label': a, 'value': a} for a in sorted(df[age_col].dropna().unique())],
                id='age_filter',
                placeholder='Select Age Group',
                multi=True
            ),
        ], style={'width': '40%', 'display': 'inline-block', 'padding': '10px'}),
    ], style={'textAlign': 'center'}),

    html.Hr(),

    # Graphs
```

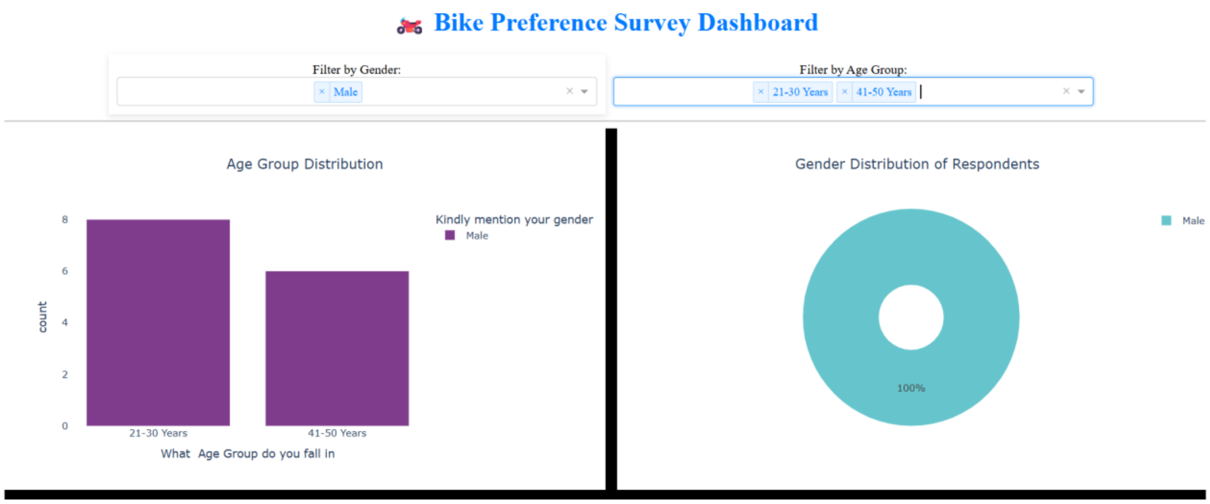
```
    html.Div([
        dcc.Graph(id='age_distribution'),
        dcc.Graph(id='gender_distribution'),
        dcc.Graph(id='bike_type_preference'),
        dcc.Graph(id='age_vs_type'),
        dcc.Graph(id='brand_preference'),
        dcc.Graph(id='purpose_usage'),
    ], style={'display': 'grid','background': 'black', 'gridTemplateColumns': '1fr 1fr', 'gap':
'15px'})
])

# Interactive Callback
@app.callback(
    [Output('age_distribution', 'figure'),
     Output('gender_distribution', 'figure'),
     Output('bike_type_preference', 'figure'),
     Output('age_vs_type', 'figure'),
     Output('brand_preference', 'figure'),
     Output('purpose_usage', 'figure')],
    [Input('gender_filter', 'value'),
     Input('age_filter', 'value')]
)
def update_graphs(selected_genders, selected_ages):
    dff = df.copy()

    # Apply filters
    if selected_genders:
        dff = dff[dff[gender_col].isin(selected_genders)]
    if selected_ages:
        dff = dff[dff[age_col].isin(selected_ages)]

    # 1️⃣ Age Group Distribution
    fig1 = px.histogram(
```

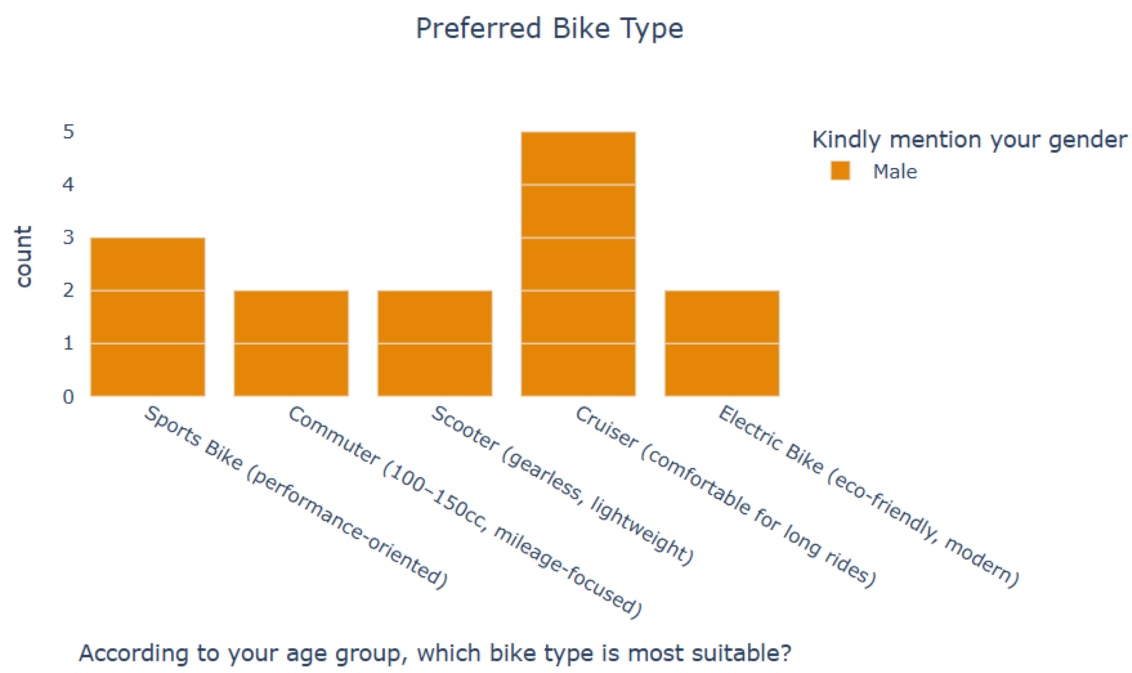
```
    dff, x=age_col, color=gender_col,
    title="Age Group Distribution",
    color_discrete_sequence=px.colors.qualitative.Bold
)

# 2️⃣ Gender Distribution
fig2 = px.pie(
    dff, names=gender_col,
    title="Gender Distribution of Respondents",
    hole=0.3,
    color_discrete_sequence=px.colors.qualitative.Pastel
)

# 3️⃣ Preferred Bike Type
fig3 = px.bar(
    dff, x=pref_col, color=gender_col,
    title="Preferred Bike Type",
    color_discrete_sequence=px.colors.qualitative.Vivid
)

# 4️⃣ Age Group vs Bike Type
fig4 = px.histogram(
    dff, x=age_col, color=pref_col,
    title="Age Group vs Preferred Bike Type",
    barmode='group',
    color_discrete_sequence=px.colors.qualitative.Dark24
)
```

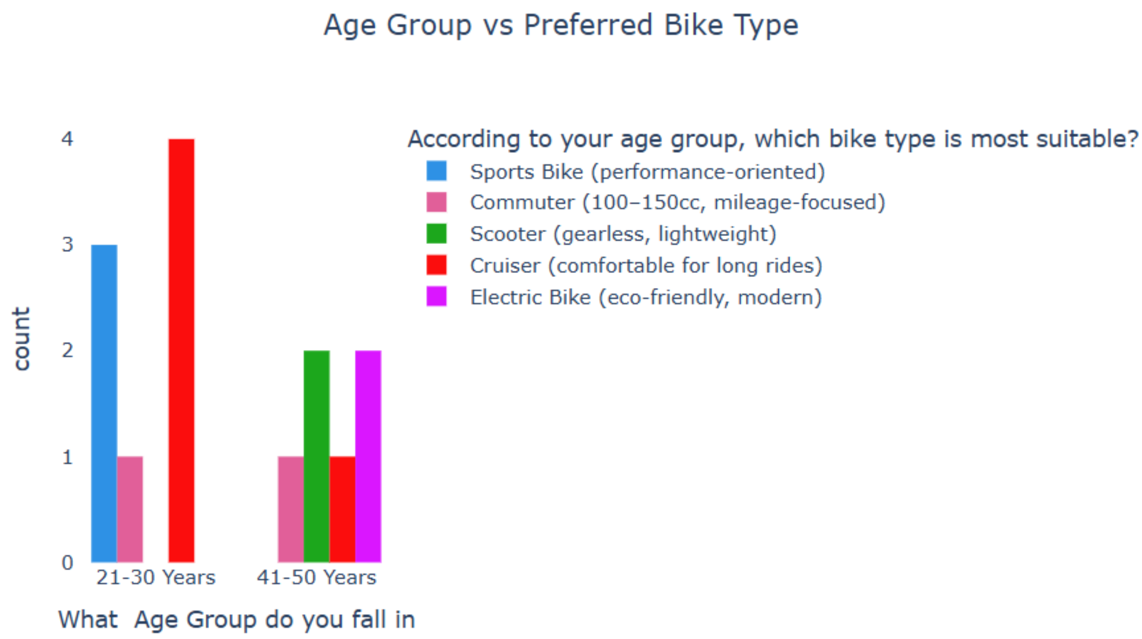
```
    # 5️⃣ Feature Importance (acts as Brand Preference)
    fig5 = px.bar(
        dff, x=brand_col, color=gender_col,
        title="Feature Importance when Choosing a Bike",
        color_discrete_sequence=px.colors.sequential.Magma
    )

    # 6️⃣ Two-wheeler Ownership (acts as Purpose/Usage)
    fig6 = px.bar(
        dff, x=usage_col, color=age_col,
        title="Bike Ownership across Age Groups",
        color_discrete_sequence=px.colors.sequential.Teal
    )

    # Common formatting
    for f in [fig1, fig2, fig3, fig4, fig5, fig6]:
        f.update_layout(title_x=0.5, plot_bgcolor='white')

    return fig1, fig2, fig3, fig4, fig5, fig6

# Run the App
if __name__ == '__main__':
    app.run()
```

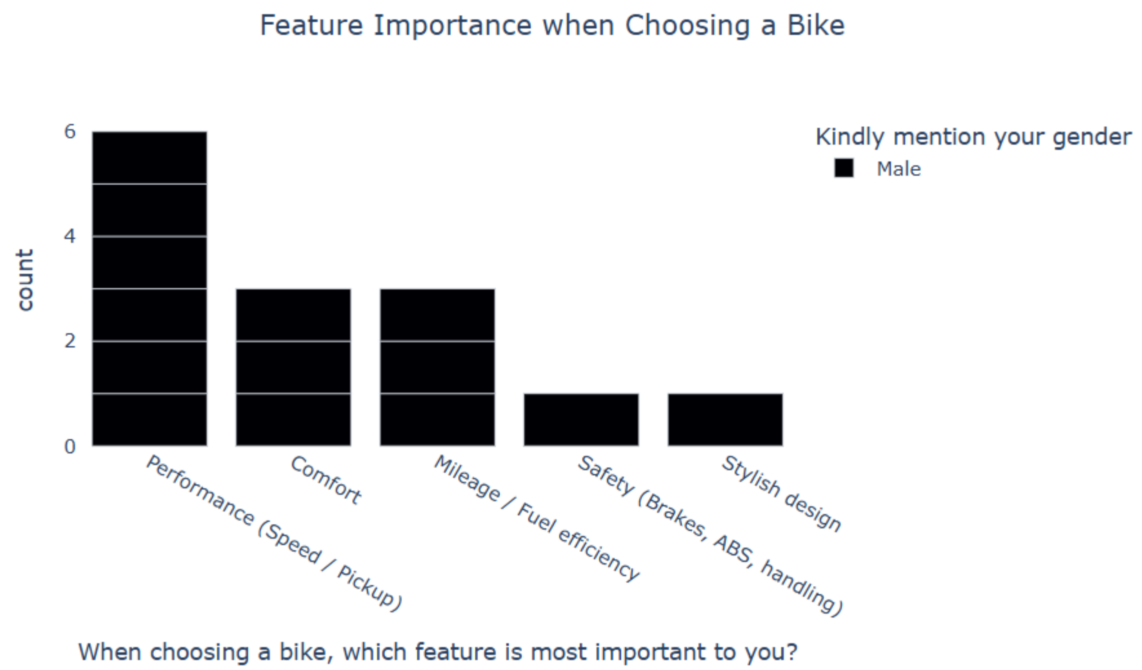
CHAPTER 9

RESULTS AND APPENDIX






Age Group vs Preferred Bike Type
count
0
1
2
3
4
21-30 Years
41-50 Years
What Age Group do you fall in
According to your age group, which bike type is most suitable?
Sports Bike (performance-oriented)
Commuter (100–150cc, mileage-focused)
Scooter (gearless, lightweight)
Cruiser (comfortable for long rides)
Electric Bike (eco-friendly, modern)


Feature Importance when Choosing a Bike
count
0
2
4
6
Performance (Speed / Pickup)
Comfort
Mileage / Fuel efficiency
Safety (Brakes, ABS, handling)
Stylish design
When choosing a bike, which feature is most important to you?
Kindly mention your gender
Male

CONCLUSION AND FUTURE SCOPE

Conclusion The primary objective of this project was to move beyond subjective opinions and establish a data-driven framework for understanding two-wheeler consumer behavior. By analyzing the relationship between demographic factors (Age and Gender) and vehicle preferences, this study successfully highlighted the distinct patterns that drive purchasing decisions in the local market.

Key takeaways from the analysis include:

- Demographic Influence: The data confirmed that age is a significant determinant of vehicle choice. While younger demographics (Student / <20 years) showed a strong inclination towards performance and aesthetics, older demographics (41+ years) consistently prioritized comfort and mileage, validating the hypothesis of a "lifecycle shift" in consumer needs.
- The Paradox of Choice: The survey results indicated that a significant portion of buyers face confusion due to conflicting advice. The dashboard successfully visualized these conflicts (e.g., the trade-off between Power and Mileage), providing a clearer picture of market reality.
- Technical Implementation: From a technical perspective, the project successfully demonstrated the power of Python and Flask in transforming raw data into actionable insights. The development of an interactive dashboard using Dash proved to be superior to static reports, as it allowed users to dynamically filter and explore the data based on their specific requirements.

In summary, this project serves as a bridge between raw consumer data and informed decision-making. It provides a valuable tool for both prospective buyers, who can see what their peers prefer, and potential market analysts, who can understand the pulse of the community.

Limitations of the Study While the project achieved its core objectives, a few limitations were identified:

- Sample Size: The dataset, while structured, represents a specific local community. A larger, more diverse dataset would be required to make nationwide generalizations.
- Static Data Source: The current dashboard operates on a static CSV file. It does not yet support real-time data ingestion where new survey responses update the graphs automatically without a server restart.

Future Scope To further enhance the utility and technical depth of this project, the following future improvements are proposed:

1. Machine Learning Integration: A predictive model (e.g., Random Forest or Decision Tree) could be trained on this dataset to build a "Recommendation Engine." This feature would ask a user for their age and budget and *predict* the best bike type for them automatically.
2. Real-Time Database Connection: Migrating from a CSV file to a cloud database (like MongoDB or SQL) would allow the dashboard to reflect new survey responses in real-time.
3. Sentiment Analysis: If the survey included more open-ended text questions (e.g., "Review your current bike"), Natural Language Processing (NLP) could be used to analyze the sentiment (positive/negative) of different bike brands.

BIBLIOGRAPHY

1. Technical Documentation & Software Libraries

These references validate the technology stack you used (Python, Flask, Dash).

- **The Pandas Development Team.** (2024). *Pandas: Python Data Analysis Library*. Retrieved from https://pandas.pydata.org/docs/
- **Plotly Technologies Inc.** (2023). *Dash Python User Guide: Interactive Data Visualization*. Retrieved from https://dash.plotly.com/
- **Plotly Express.** (2023). *High-level interface for data visualization*. Retrieved from https://plotly.com/python/plotly-express/
- **Pallets Projects.** (2024). *Flask Documentation (Version 3.0)*. Retrieved from https://flask.palletsprojects.com/

2. Web Resources & Tutorials

- **W3Schools.** (2023). *Python Tutorial and Data Science Modules*. Retrieved from https://www.w3schools.com/python/
- **Towards Data Science.** (2022). *Building a Dashboard with Dash and Plotly*. Medium.
- **Stack Overflow.** (2023). *Community discussions on Pandas DataFrames and Error Handling*.

ACKNOWLEDGEMENT

The successful completion of this project, **"Analysis of Bike Preferences Survey using Python & Dash,"** would not have been possible without the guidance and support of many respected individuals.

First and foremost, I would like to express my sincere gratitude to my project guide, **[Dr. Kumar B I D]**, for their invaluable mentorship, continuous encouragement, and technical advice throughout the development of this application. Their insights helped me navigate the complexities of data analysis and web deployment.

I am also deeply grateful to our Head of Department, **[Dr. Kumar B I D]**, and our Principal, **[Dr. Santosh M Muranal]**, for providing the necessary academic resources and a conducive environment for learning at **[Amruta Institute of Engineering and Management Sciences]**.

A special thanks goes to the **respondents of my survey**. This project relies entirely on primary data, and I am grateful to every individual who took the time to fill out the Google Form. Their honest responses regarding vehicle choices and preferences form the backbone of this analysis.

I would also like to thank my friends and classmates for their constructive criticism and help during the testing phase of the dashboard.

Finally, I express my gratitude to my parents and family for their unwavering support and patience during the course of this study.

[Simon Leo Alexander] [1AR23IS042]