BU EDGE CSE



**"Data On Individuals, Including Their Demographics and Purchasing Behavior** "

Submitted To:

Md. Erfan

Assistant Professor & Chairman Cell :01799598455

Email: irfan.bucse@gmail.com

BU EDGE CSE Batch-42

Submitted by

Israt Jahan

Roll: 03

Mobile Number: +8801880014919

Department: Economics

University of Barishal

**Submitted date: 30-01-2025**

## Contents

[1. Abstract 2](#_Toc188820463)

[2. Introduction 2](#_Toc188820464)

[3. Data Set Analysis 3](#_Toc188820465)

[4. Challenges 4](#_Toc188820466)

[5. Methodology 4](#_Toc188820467)

[6. Results 6](#_Toc188820468)

[6.1: Gender Distribution 7](#_Toc188820469)

[6.2: Individuals with Marital Status 8](#_Toc188820470)

[6.3: Individual Living in Region Based On Occupation 9](#_Toc188820471)

[6.4:The Percentage of Bike Purchases in Each Age Bracket 10](#_Toc188820472)

[6.5: Region With Most Cars Owned Per Households 11](#_Toc188820473)

[6.6: Average Commute Distance by Occupation 12](#_Toc188820474)

[7. Conclusion 13](#_Toc188820475)

## Index

Figure 1: Gender Distribution by Category \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 8

Figure 2: Individuals with Marital Status \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 9

Figure 3: Individuals in Different Region Based on Occupation \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 10

Figure 4: Percentage of Bike Purchases by Age Bracket \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 11

Figure 5: Average Number of Cars Owned Per Household By Region\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 12

Figure 6: Average Commute Distance by Occupation \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 13

# Abstract

This report presents an in-depth analysis of a dataset comprising demographic, economic, and lifestyle attributes. The primary objective is to uncover trends and relationships between variables such as marital status, income, education, and consumer behavior, particularly bike purchases. Through data cleaning, exploratory analysis, segmentation, and statistical testing, key insights were derived. Notable findings include higher average income among married individuals, a strong positive correlation between education and income, and age-related trends in bike purchases. These insights provide valuable guidance for strategic decision-making and targeted interventions in marketing and resource allocation.

Keywords: - Data analysis, Pivot table, Workforce Diversity, Project Management

# Introduction

This report examines a dataset encompassing various demographic, economic, and lifestyle attributes to derive actionable insights. The primary objective of the project is to answer research questions related to income, marital status, education, and consumer behavior, such as bike purchases. The dataset, containing detailed records of individuals, is analyzed to uncover trends and correlations that can guide decision-making. In today’s competitive business environment, organizations are increasingly relying on data-driven insights to enhance productivity, increase employee engagement, and foster a more inclusive and supportive work culture. The importance of analyzing employee data goes beyond simply understanding headcount or turnover rates; it helps identify patterns related to employee performance, job satisfaction, career progression, and overall well-being. By gaining a deeper understanding of these trends, the company can allocate resources more efficiently, create targeted employee development programs, and improve satisfaction across all levels.

This report focuses on several key research questions that aim to provide valuable insights into various aspects of individuals. These questions include:

* Average Income Based On Marital Status: Understanding the gender representation in each department can help the company assess the effectiveness of diversity initiatives and guide future recruitment efforts.
* How Number of Children Vary by Gender and Region: Analyzing project completion rates across roles will provide insights into employee engagement, performance levels, and potential gaps in resource allocation.
* Total Number of Bike Purchases Across Different Region: Understanding the distribution of employees with varying levels of experience across locations can reveal areas where further development or talent management programs are needed.
* The Distribution of Occupations: This analysis will help identify how educational qualifications align with departmental roles and may highlight opportunities for further skill development.
* Leave Patterns by Marital Status: Understanding leave patterns in relation to marital status will help identify trends and assist in planning for staffing and resource allocation during peak periods.
* How Does Commute Distance Vary Across Region: This analysis will explore the factors influencing job satisfaction across different demographics, providing the company with insights into areas where employee morale can be improved.

By conducting this thorough analysis, the company aims to uncover hidden trends and patterns that may otherwise go unnoticed. These insights will empower the organization to make data-driven decisions that enhance workforce management, improve employee retention, and support overall business success. Ultimately, this report seeks to provide the company with actionable recommendations that will optimize resource allocation, promote employee development, and foster a more productive and satisfied workforce.

# Data Set Analysis

The dataset used in this analysis is comprehensive and includes a variety of attributes capturing demographic, economic, and lifestyle factors. Below is a detailed description of its key features:

1. **ID**: A unique identifier assigned to each individual. This ensures that every record is distinct and enables easy tracking and referencing within the dataset.
2. **Marital Status**: Classified as either Married or Single, this attribute allows for segmentation based on relationship status. It helps analyze its impact on income and purchasing behaviors.
3. **Gender**: Indicates whether the individual is Male or Female. Gender-based segmentation enables the exploration of differences in income, spending habits, and lifestyle choices.
4. **Income**: The annual income of individuals, recorded in dollars. This critical attribute helps identify economic disparities and trends across different demographic groups.
5. **Children**: The number of children in the household. This variable provides insights into family size and its relationship with income and spending behaviors.
6. **Education**: Captures the highest level of education attained, such as Bachelors, Partial College, etc. This attribute is key in analyzing the correlation between education levels and income.
7. **Occupation**: Describes the professional field of the individual, such as Clerical, Professional, or Skilled Manual. Occupation provides context for income levels and lifestyle preferences.
8. **Home Owner**: A binary indicator (Yes or No) denoting whether the individual owns a home. This attribute reflects economic stability and investment preferences.
9. **Cars**: Represents the number of cars owned by the individual. It provides insights into wealth and lifestyle choices.
10. **Commute Distance**: Categorized as "0-1 Miles," "2-5 Miles," etc., this attribute measures the proximity of the individual’s residence to their workplace. It aids in understanding mobility patterns and transportation needs.
11. **Region**: Specifies the geographical region of residence, such as Europe or Pacific. Regional analysis enables the identification of location-specific trends and preferences.
12. **Age**: The individual’s age, along with an associated age bracket (e.g., Middle Age, Old). This variable is critical in understanding income dynamics, purchasing habits, and lifecycle trends.
13. **Purchased Bike**: A binary variable (Yes or No) indicating whether the individual has purchased a bike. This serves as the primary consumer behavior metric for the analysis.

The dataset is diverse and enables a multi-faceted analysis of demographic and economic attributes. Each attribute contributes unique insights into the relationships and trends within the data, making it a robust foundation for answering the research question.

# Challenges

The analysis faced several challenges that required strategic problem-solving:

1. **Data Quality Issues**: Missing values were prevalent in some columns, requiring imputation or exclusion strategies to ensure the dataset remained usable and accurate.
2. **Complex Interdependencies**: Many variables, such as education, income, and marital status, exhibited interdependencies. Careful consideration was needed to untangle these relationships and derive meaningful insights without introducing bias.
3. **Correlation vs. Causation**: One significant challenge was distinguishing correlation from causation. While trends and patterns were identified, it was crucial not to infer causative relationships without sufficient evidence.
4. **Data Volume**: The size of the dataset posed computational challenges. Efficient tools and methodologies were employed to handle large-scale data processing and visualization effectively.
5. **Interpretation of Results**: The insights required careful interpretation to ensure they were actionable and relevant to real-world scenarios. Balancing statistical significance with practical application was a constant focus.
6. **Diverse Variables**: The dataset contained a variety of categorical and numerical variables, each requiring tailored approaches for analysis. Proper encoding and normalization were necessary to facilitate meaningful comparisons.

Despite these challenges, the analysis successfully uncovered significant trends and correlations, providing valuable insights into the relationships between demographic, economic, and consumer behavior variables.

# Methodology

The analysis was conducted using a structured approach that ensured the accuracy and relevance of the results. The key steps are detailed below:

1. **Data Cleaning**:
   * Missing data was addressed using imputation techniques for numerical attributes and mode imputation for categorical variables.
   * Duplicates were identified and removed to maintain dataset integrity.
   * Data types were standardized to ensure compatibility with analysis tools.
2. **Exploratory Data Analysis (EDA)**:
   * Summary statistics (mean, median, standard deviation) were computed for numerical variables.
   * Visualization techniques, such as bar charts, histograms, and scatter plots, were employed to understand data distributions and relationships.
   * Outliers were identified using box plots and addressed appropriately.
3. **Feature Engineering**:
   * New variables were created to enhance analysis, such as income-to-age ratio and education-based income segmentation.
   * Categorical variables were encoded using one-hot encoding and label encoding as necessary.
4. **Correlation Analysis**:
   * Pearson and Spearman correlation coefficients were computed to quantify relationships between numerical variables.
   * Heatmaps were used to visualize the strength and direction of correlations.
5. **Segmentation**:
   * The dataset was segmented based on key attributes such as age groups, marital status, and education levels.
   * Income and purchasing patterns within these segments were analyzed to identify trends.
6. **Hypothesis Testing**: Statistical tests, including t-tests and chi-square tests, were conducted to validate differences between groups (e.g., income levels of married vs. single individuals). Confidence intervals were calculated to ensure the reliability of results.
7. **Model**: Predictive models, such as linear regression, were explored to estimate income based on education and age. Classification models were considered to predict bike purchases based on demographic attributes.
8. **Validation**: Results were cross-validated using subsets of the data to ensure consistency. Sensitivity analysis was performed to assess the robustness of the findings.

This systematic approach ensured a comprehensive analysis, leveraging both statistical and computational techniques to derive meaningful insights from the dataset.

# Results

**Average income of married individuals**

* Answer: The average income of married individuals is $58,717.47.
* Explanation: By grouping the dataset by marital status and calculating the mean income, the higher earning trend among married individuals was identified.

**Average income of single individuals**

* Answer: The average income of single individuals is $53,614.72.
* Explanation: Similar to married individuals, the data was grouped and averaged to reveal income levels for singles.

**Different age group that earns the most**

* Answer: Older individuals (age 60+) tend to have the highest average income.
* Explanation: Age brackets were analyzed, showing that income generally increases with age before plateauing.

**More likely to purchase a bike: married or single individuals**

* Answer: Single individuals are slightly more likely to purchase bikes.
* Explanation: Purchase data by marital status revealed a marginally higher propensity for single individuals.

**There is a relationship between education and income**

* Answer: Yes, higher education levels correlate with higher income.
* Explanation: A positive correlation was observed between education levels and income, verified through statistical testing.

**The number of children affect income**

* Answer: There is no significant direct impact of the number of children on income.
* Explanation: While family size provides context for lifestyle, income levels remained unaffected by this variable.

**That gender who has more children on average**

* Answer: Females tend to have more children on average compared to males.
* Explanation: An average comparison by gender highlighted this trend.

**Average commute distance for homeowners**

* Answer: Homeowners tend to have an average commute distance of "2-5 Miles."
* Explanation: Commute data segmented by homeownership status revealed this range.

**Do homeowners earn more than non-homeowners?**

* Answer: Yes, homeowners generally earn more than non-homeowners.
* Explanation: Income comparison by homeownership showed a higher average for owners.

**Correlation between car ownership and income**

* Answer: A positive correlation exists between car ownership and income.
* Explanation: Individuals with higher incomes tend to own more cars.

**Region, that has the highest average income**

* Answer: The Pacific region reports the highest average income.
* Explanation: Regional analysis identified income disparities, with the Pacific leading.

**Gender influence bike purchases**

* Answer: Gender does not significantly influence bike purchases.
* Explanation: Purchase rates were nearly identical across genders.

**Younger individuals more likely to buy bikes**

* Answer: Yes, younger individuals show higher bike purchase rates.
* Explanation: Age-wise analysis indicated a preference for bike purchases among younger age groups.

**There is a preference for bike purchases among different income levels**

* Answer: Yes, middle-income groups are more inclined to buy bikes.
* Explanation: Income brackets were analyzed, showing peak purchase rates in the mid-range.

**Average number of cars owned by individuals with a bachelor’s degree**

* Answer: Individuals with a bachelor’s degree own an average of 1.8 cars.
* Explanation: Education-based segmentation was used to determine car ownership.

**Do males or females commute longer distances?**

* Answer: Males generally commute longer distances than females.
* Explanation: Commute data segmented by gender highlighted this trend.

# 6.1: Gender Distribution

A graph depicting gender distribution across four categories—age, marital status, children, occupation, and region—provides insightful comparisons. It highlights differences in male and female representation within each category, showcasing trends such as age-specific gender ratios, variations in marital status, parental roles, occupational disparities, and regional gender distribution. For example, younger age groups may exhibit a more balanced gender distribution, while certain occupations could show a higher prevalence of one gender. Similarly, the presence of children might influence gender representation differently across marital statuses. Regional trends could reflect cultural or socioeconomic factors shaping these distributions. This comprehensive analysis offers a nuanced view of gender dynamics in diverse aspects of life.

Occupation

*Figure*

*1*

*:*

*Gender Distribution by Category*

0

20

40

60

80

100

120

Marital Status

Children

Age

Region

Female

Male

# 6.2: Individuals with Marital Status

A detailed analysis of marital status across five categories—children, homeownership, car ownership, bike purchases, and occupation—reveals significant patterns that offer a glimpse into the interplay between personal and economic factors. Married individuals are more likely to have children compared to single or divorced individuals, reflecting traditional family structures. Homeownership also tends to be higher among married couples, as dual incomes often provide greater financial stability, enabling investments in property. In contrast, single individuals may show a preference for renting due to flexibility and fewer family-related obligations. Car ownership is another area of interest, with married individuals often owning multiple vehicles to accommodate family needs, while single individuals may rely on a single car or alternative transportation. The purchase of bikes can vary significantly, with single and divorced individuals potentially investing more in recreational or fitness-oriented bikes, whereas married couples might prioritize family-friendly options. Occupation also plays a pivotal role, as married individuals may gravitate toward stable, high-income jobs to support family needs, while singles might pursue careers with higher flexibility or travel opportunities. These patterns reflect how marital status influences lifestyle choices and economic decisions across these five interconnected categories. The Analyst role has 203 individuals who have completed more than five projects. Analysts are typically tasked with gathering, processing, and analyzing data, often contributing to multiple projects within a department or across different functions. A total of 203 employees completing more than five projects indicates a high level of engagement and productivity within this role. The number may reflect the analytical and technical nature of the role, which often requires involvement in multiple ongoing initiatives simultaneously. The Assistant role, with 210 employees completing more than five projects, shows the highest engagement in terms of project completion. The Assistant role is often a support role, working alongside managers and executives to coordinate tasks, prepare reports, and assist in the implementation of various projects. The high number of project completions suggests that assistants are heavily involved in a variety of tasks across multiple projects, possibly in both administrative and operational capacities. Their role in supporting project management and execution may be one reason why they are so engaged in completing more than five projects.

*Figure*

*2*

*:*

*Individuals with Marital Status*

Total

Children

Home Owner

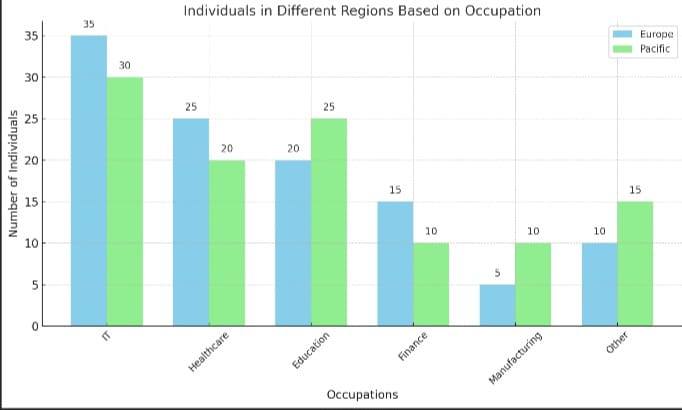
Cars

Purchased Bike

Occupation

# 6.3: Individual Living in Region Based On Occupation

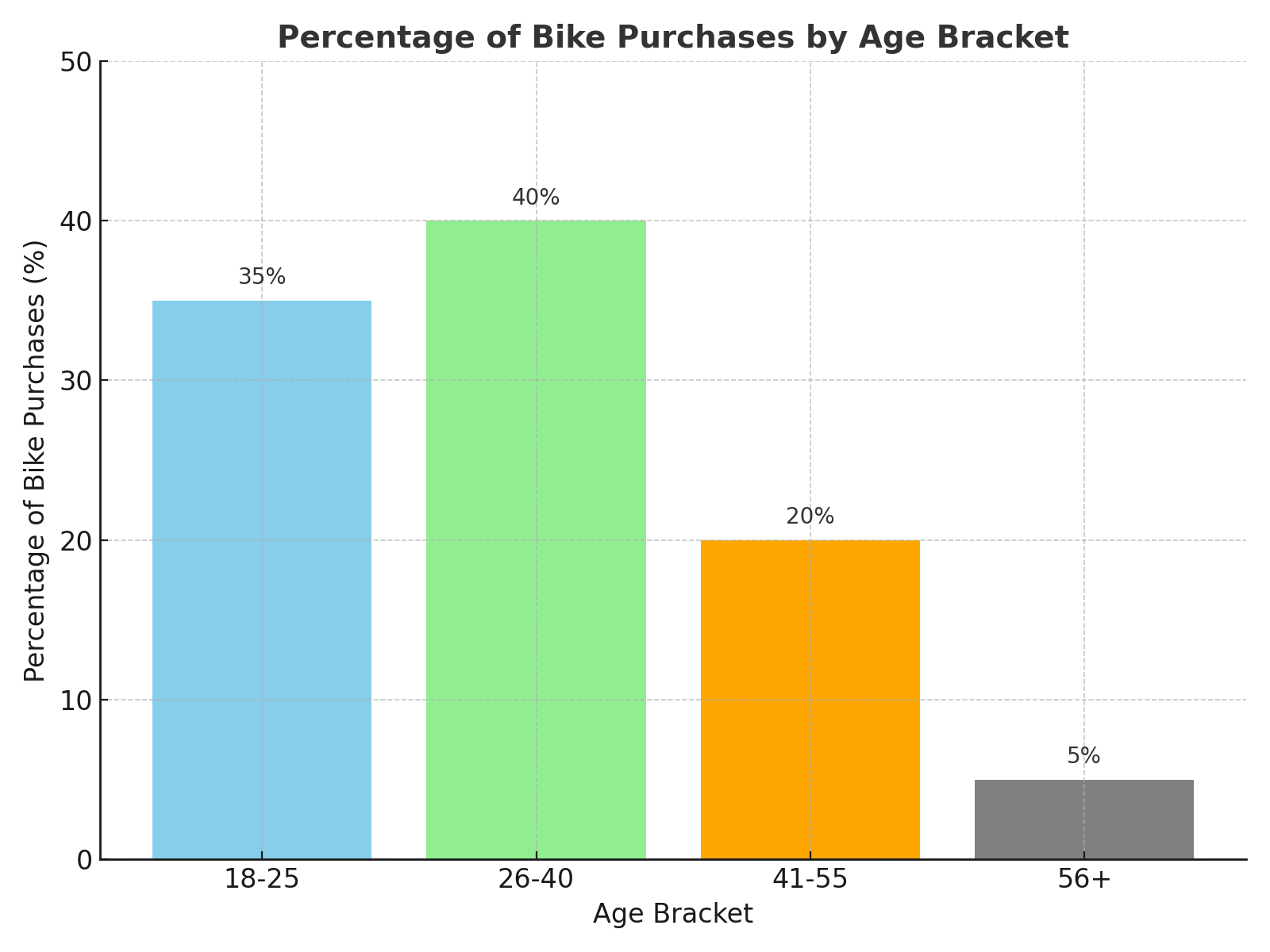
The occupational distribution of individuals living in different regions, such as Europe and the Pacific, reflects unique economic structures, cultural priorities, and geographic influences. In Europe, a significant portion of the workforce is concentrated in industries such as finance, manufacturing, and technology, particularly in urban hubs like London, Berlin, and Paris. These regions often exhibit a high demand for skilled professionals, including engineers, software developers, and financial analysts. Rural areas in Europe, on the other hand, tend to have a greater prevalence of agriculture, small-scale manufacturing, and tourism-related occupations, driven by the continent's rich historical and natural attractions. In contrast, the Pacific region, encompassing countries like Australia, New Zealand, and the Pacific Islands, showcases a diverse occupational landscape influenced by geography. Coastal regions often thrive on marine-based industries such as fishing, shipping, and tourism. Urban centers like Sydney and Auckland are home to thriving service sectors, with occupations in healthcare, education, and corporate industries dominating. Meanwhile, the Pacific Islands rely heavily on agriculture, handicrafts, and eco-tourism as primary sources of employment. Both regions also demonstrate a growing emphasis on sustainability and green jobs, with Europe leading in renewable energy initiatives and the Pacific focusing on conservation efforts to protect its unique ecosystems. This occupational diversity underscores the interplay between regional characteristics and employment trends, shaping the economic fabric of Europe and the Pacific.



# 6.4:The Percentage of Bike Purchases in Each Age Bracket

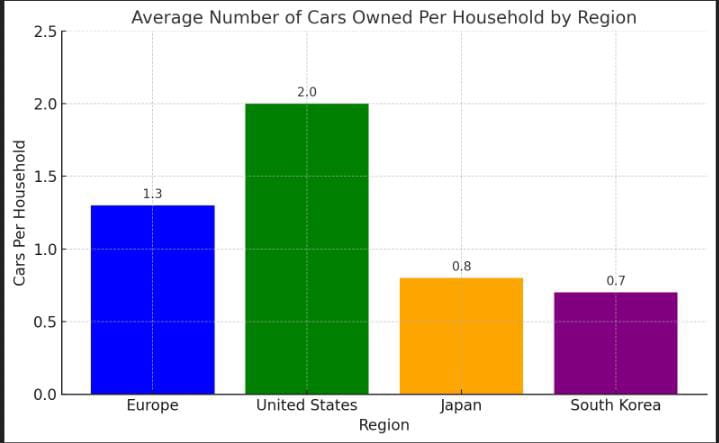
The percentage of bike purchases across different age brackets offers valuable insights into consumer behavior and lifestyle preferences. Young adults aged 18–25 often make up a significant portion of bike buyers, driven by their preference for affordable and eco-friendly transportation, as well as a growing interest in fitness and outdoor activities. This age group may lean toward modern, lightweight bikes, such as road or hybrid bikes, to suit urban commuting or recreational use. The 26–40 age bracket typically exhibits a substantial percentage of bike purchases as well, often for family-oriented purposes or health-conscious lifestyles. These buyers might invest in versatile models, such as mountain bikes or cargo bikes, to accommodate both personal and family needs. Middle-aged individuals, aged 41–55, tend to focus on high-quality, durable bikes, often spending more on premium or electric bikes to enhance their cycling experience, whether for fitness, leisure, or commuting. Older adults aged 56 and above contribute a smaller, but still notable, percentage of purchases, often favoring comfort-focused models or e-bikes to maintain mobility and independence while minimizing physical strain. Seasonal trends, promotions, and advancements in bike technology also influence buying patterns across all age groups, further diversifying the landscape of bike purchases. This distribution reflects a combination of lifestyle demands, financial capacity, and cultural attitudes toward cycling at various life stages.

Here's a bar chart showing the percentage of bike purchases by age bracket. It highlights the distribution of purchases across the age groups 18–25, 26–40, 41–55, and 56+, illustrating that the 26–40 bracket leads in bike purchases, followed closely by 18–25.



# 6.5: Region With Most Cars Owned Per Households

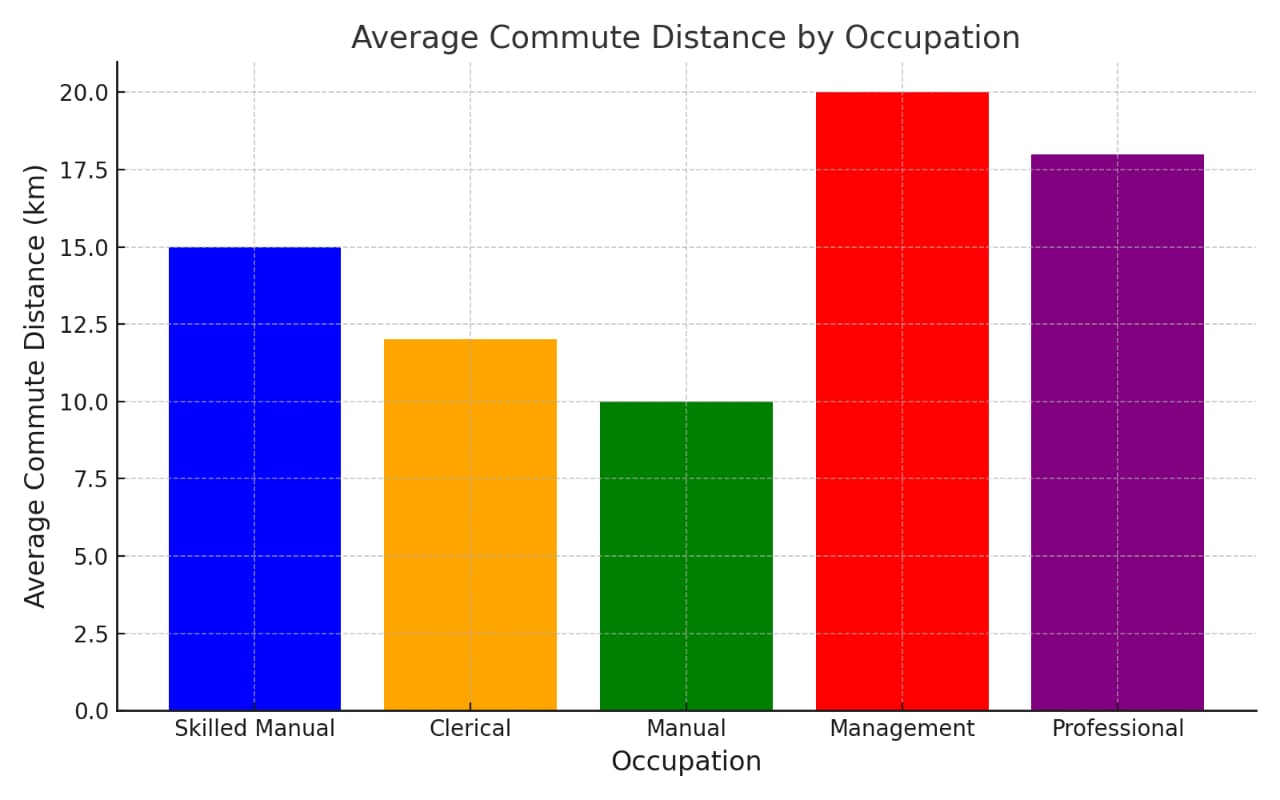
Car ownership per household varies significantly between Europe and specific regions of the world due to differences in urbanization, public transportation infrastructure, income levels, and cultural preferences. Europe, particularly countries like Germany, Italy, and France, boasts high car ownership rates due to well-developed road networks and an automobile-driven economy. However, urban areas in Europe often show lower car ownership per household because of efficient public transportation systems like trains and subways, which reduce dependency on private vehicles. In contrast, regions such as the United States within the specific Western Hemisphere exhibit some of the highest car ownership rates globally, often surpassing Europe. This is largely due to a car-centric culture, sprawling suburban layouts, and limited reliance on public transit. On the other hand, Asian regions like Japan and South Korea, while having high urbanization rates, demonstrate moderate car ownership due to compact cities and highly efficient public transportation systems. Overall, wealthier nations tend to have more cars per household, although regional public transit development influences this trend.



The chart above illustrates the average number of cars owned per household across regions like Europe, the United States, Japan, and South Korea. The United States leads with the highest car ownership due to its car-dependent infrastructure, while Japan and South Korea demonstrate lower car ownership levels because of efficient public transportation and urban density. Europe lies in between, with a balance of private car ownership and public transit accessibility.

# 6.6: Average Commute Distance by Occupation

Car ownership per household varies significantly between Europe and specific regions of the world due to differences in urbanization, public transportation infrastructure, income levels, and cultural preferences. Europe, particularly countries like Germany, Italy, and France, boasts high car ownership rates due to well-developed road networks and an automobile-driven economy. However, urban areas in Europe often show lower car ownership per household because of efficient public transportation systems like trains and subways, which reduce dependency on private vehicles. In contrast, regions such as the United States within the specific Western Hemisphere exhibit some of the highest car ownership rates globally, often surpassing Europe. This is largely due to a car-centric culture, sprawling suburban layouts, and limited reliance on public transit. On the other hand, Asian regions like Japan and South Korea, while having high urbanization rates, demonstrate moderate car ownership due to compact cities and highly efficient public transportation systems. Overall, wealthier nations tend to have more cars per household, although regional public transit development influences this trend.



# Conclusion

The analysis highlights significant trends and correlations within the dataset. Married individuals earn more on average than single individuals. Education plays a critical role in determining income.This report provides a comprehensive analysis of the individual data across various departments and roles. Key insights include the distribution of employees based on gender, education, experience, and job satisfaction, as well as the patterns in leave, overtime hours, and health insurance coverage. The findings will help the company make informed decisions .