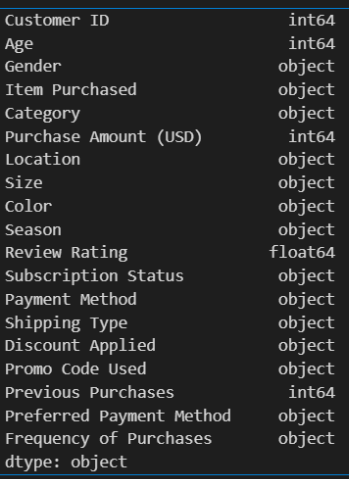
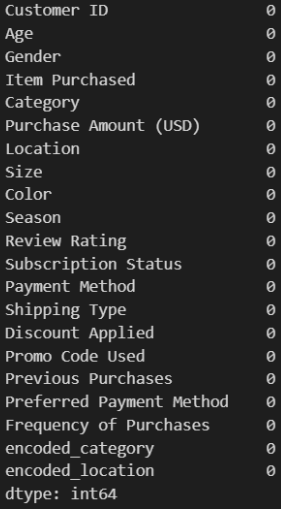
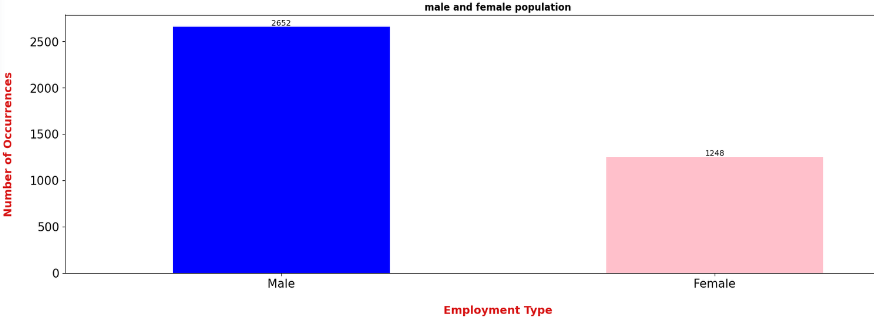
**Importing the Dataset**

****

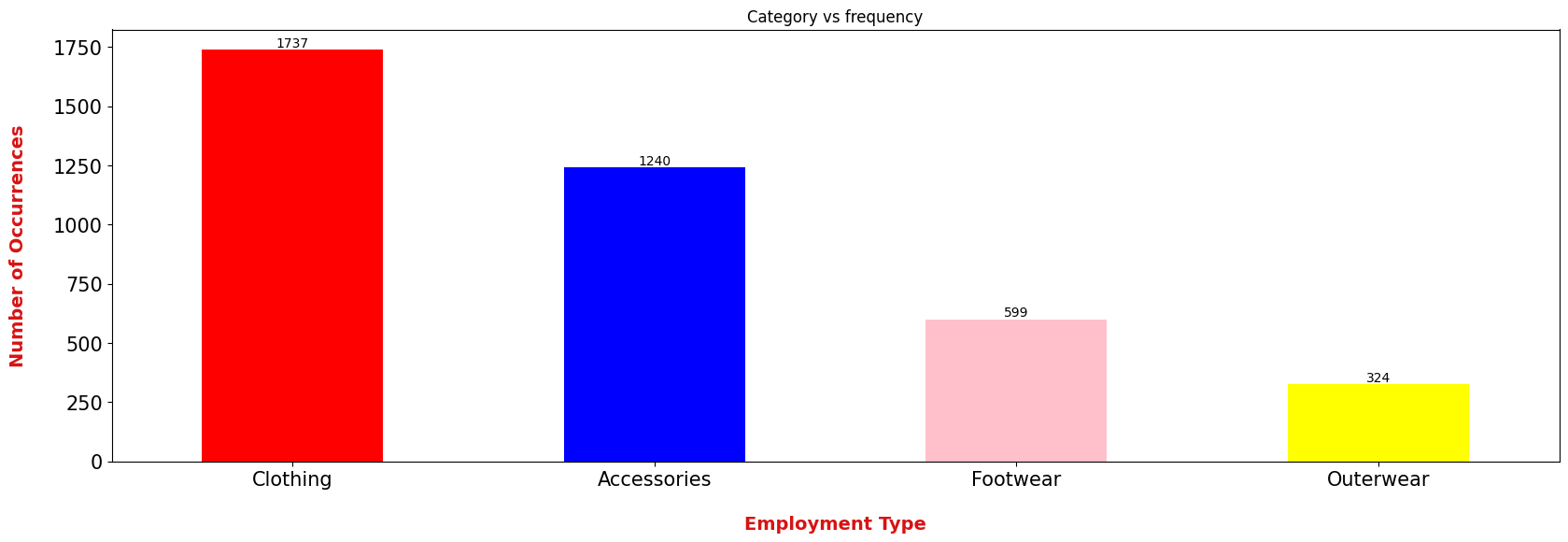
In our dataset, there aren't any missing values, and we've got 22 columns, each holding different types of information like numbers, words, and decimals. Now, for our analysis, we're narrowing down our focus to just three columns that seem most important. This helps us concentrate on specific data that we believe will give us the most useful insights. By selecting these three columns, we're aiming to dig deeper into the information they hold and uncover meaningful patterns or trends that can guide our analysis and decision-making process.

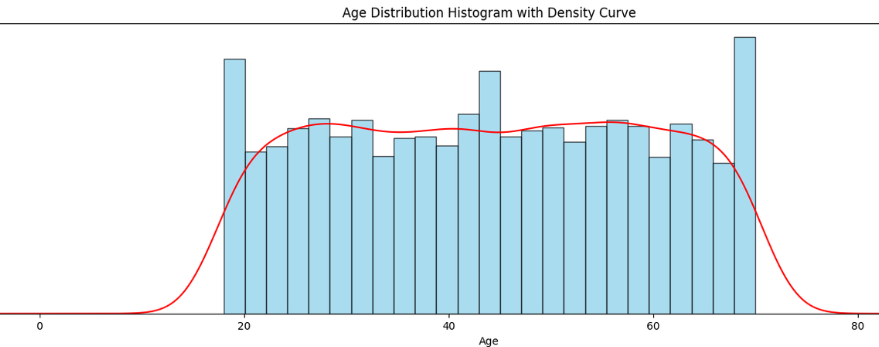
**EDA**

****

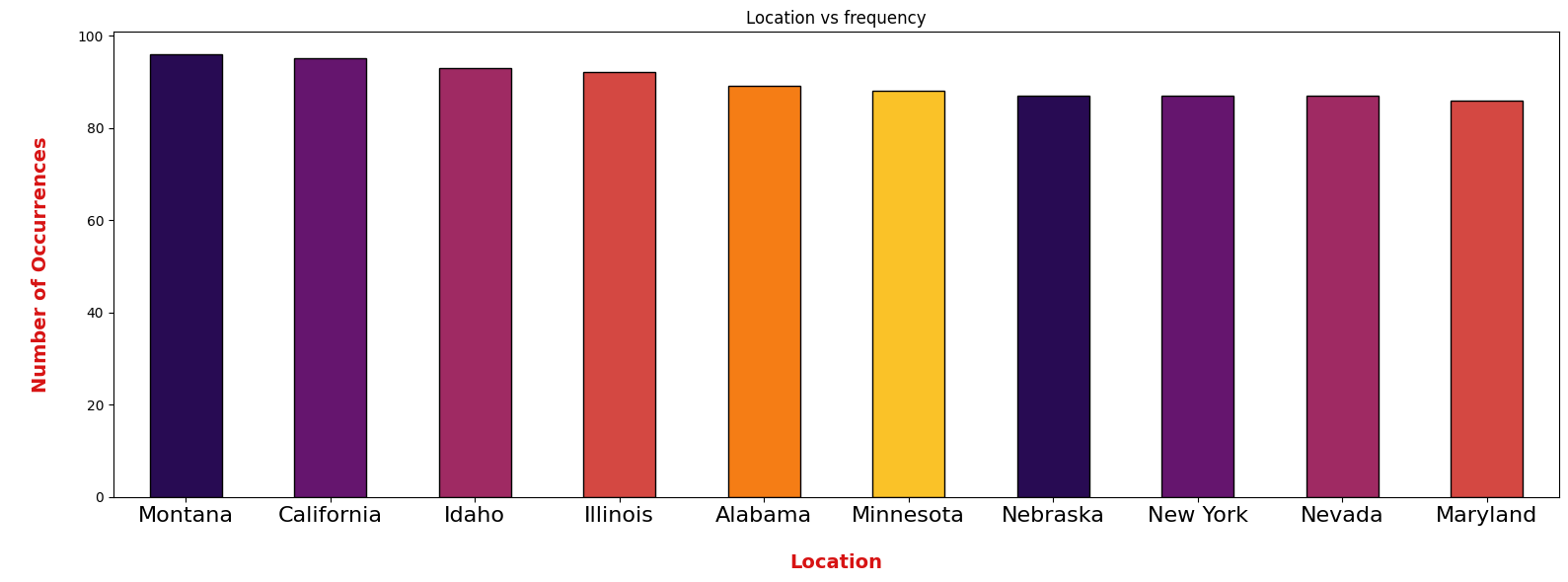
This bar graph shows the number of customers based on their gender. On the horizontal line (x-axis), you'll see the genders - males and females. The vertical line (y-axis) represents how many customers belong to each gender.

Looking at the graph, it's clear that there are more males than females in this customer dataset. Males have a higher frequency compared to females. The bars for males are taller, showing a larger number of male customers, while the bars for females are shorter, indicating a smaller number of female customers in this data set

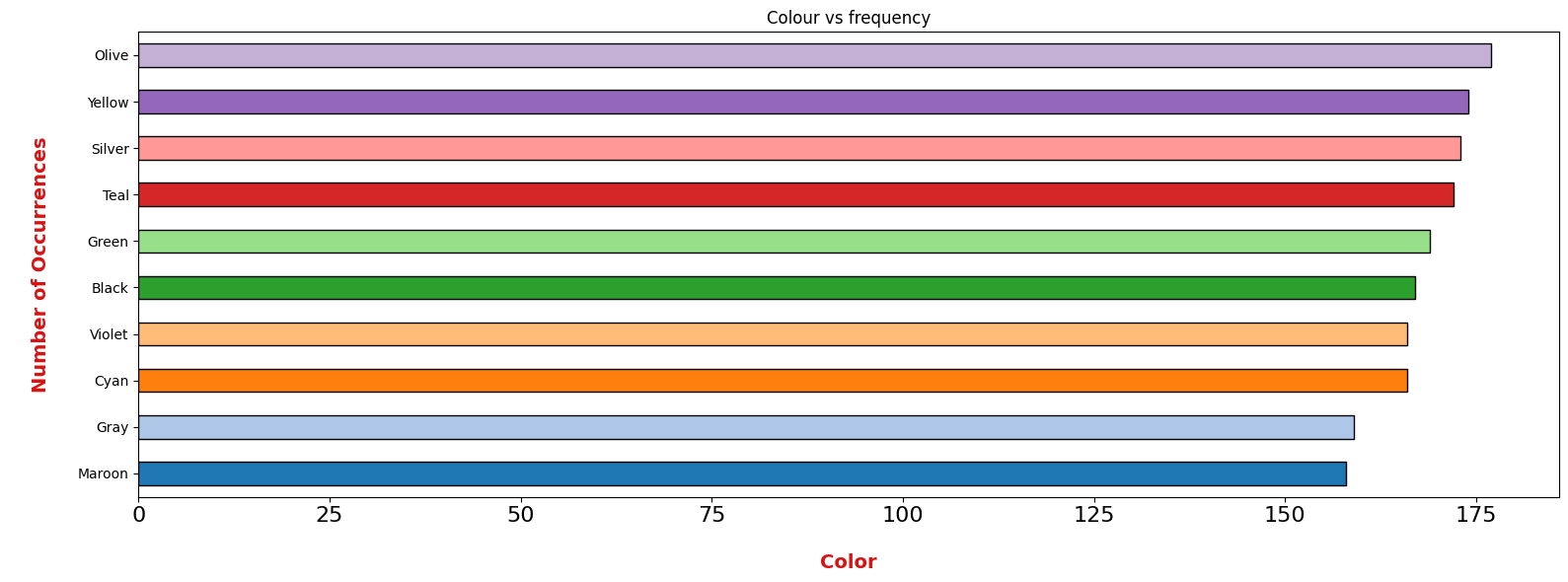
.

In the graph, we see a bar plot depicting how often customers buy products in different categories. The x-axis shows the product categories, and the y-axis represents the buying frequency. Notably, clothing emerges as the most popular category, with the highest number of purchases. On the other hand, outerwear stands out as the least frequently bought item. This simple visualization gives a quick overview of customer preferences, highlighting the strong demand for clothing and the comparatively lower interest in outerwear among the surveyed customers.****

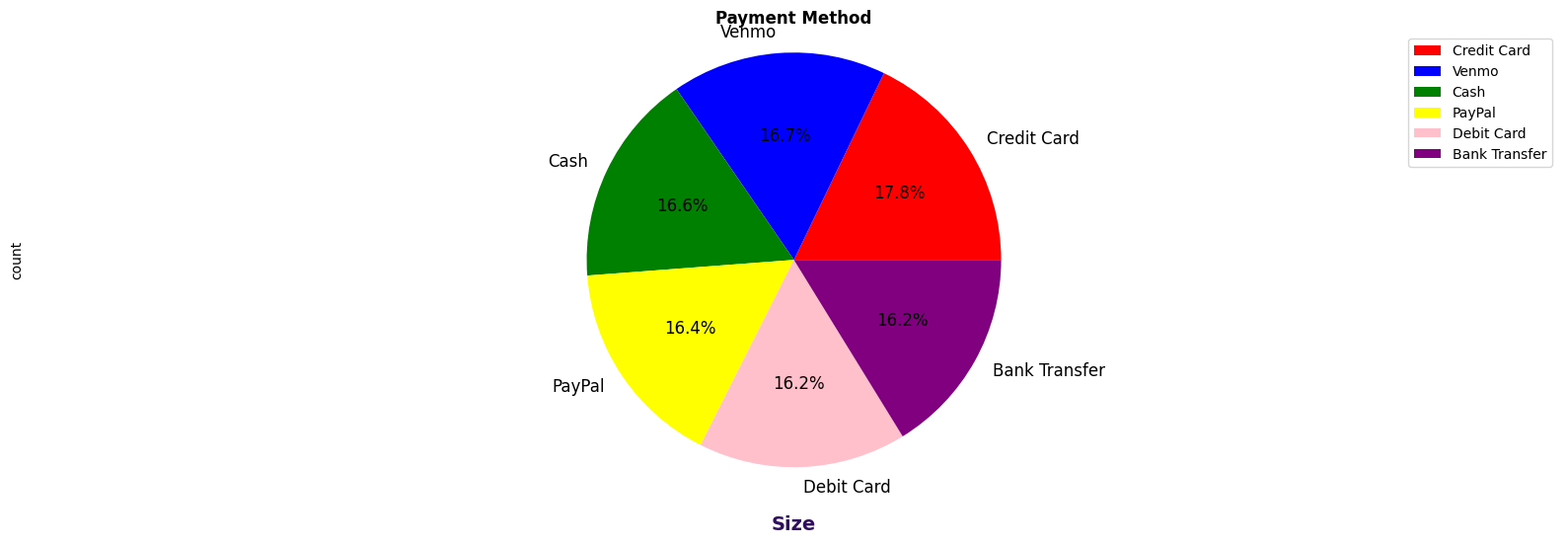
The chart displays the age distribution of customers in a dataset. It reveals that most customers fall between the ages of 20 and 70. Looking at the density curve, we observe fairly even frequencies across various age groups. Notably, ages around 20, 41, and 65 stand out as the most common. In simple terms, it shows that the dataset covers a broad age range, with a somewhat balanced representation of customers in different age brackets.



This bar graph shows where customers visit the most, with different locations on the x-axis and how often they're visited on the y-axis. Montana stands out as the most popular spot, while Maryland seems to be the least visited among them all. Interestingly, most locations have a similar number of visitors, making the graph pretty even across the board. It's like Montana's the hot spot and Maryland's the quieter one in this crowd of places customers like to go.

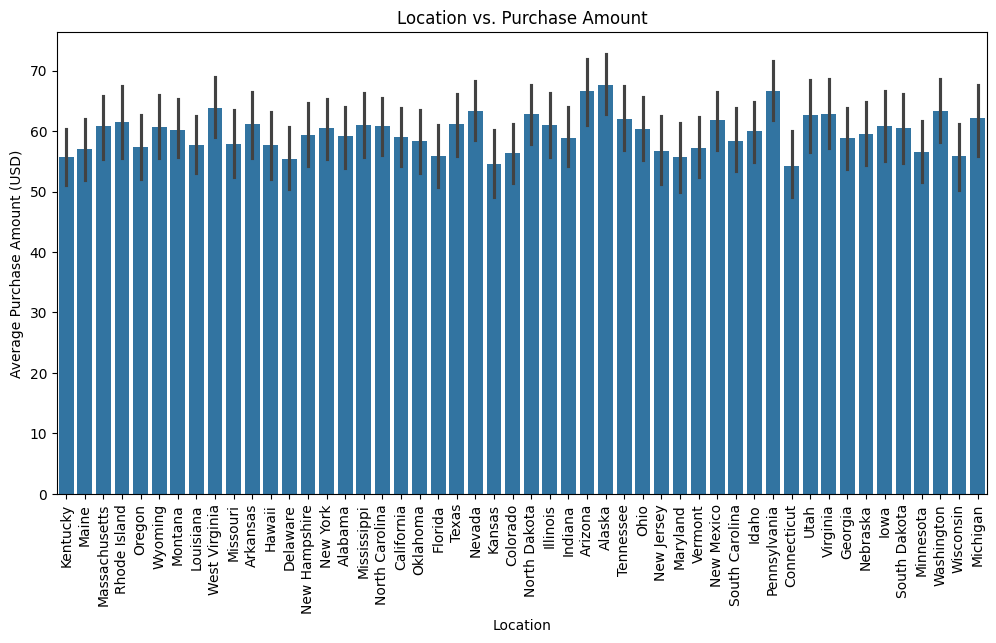


The bar graph shows which colours customers bought the most, and it's clear that olive was the top-selling colour. On the flip side, maroon was the least popular choice among buyers. This data gives a quick view of which colours flew off the shelves and which ones quite catch people's eyes.

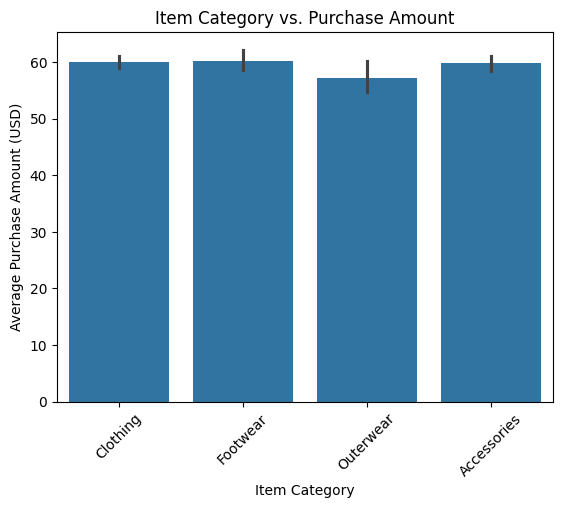


This pie chart breaks down customer preferences for payment methods. Venmo stands out as the top choice, followed closely by credit cards. On the flip side, cash is the less favoured option, accounting for only 16.6% of customer preferences. The chart provides a clear snapshot of the payment landscape, showcasing the popularity of digital options like Venmo and credit cards, while emphasizing the relatively lower preference for cash transactions. This insight can guide businesses in tailoring their payment processes to align with customer preferences and enhance overall satisfaction.

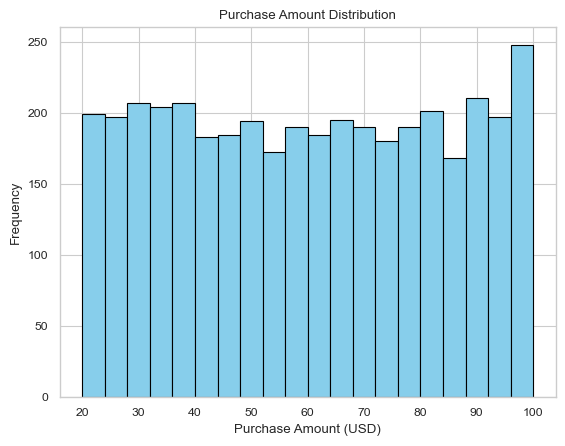


The bar graph displays how different shipping options are preferred by customers based on their gender. There are six shipping methods: express, free shipping, next day air, standard, two-day shipping, and store pickup. What stands out is that guys seem to lean towards more purchases compared to gals. It's interesting to see that across these shipping choices, males seem to have a higher preference, indicating they might be more active in selecting and utilizing these shipping methods than females. 

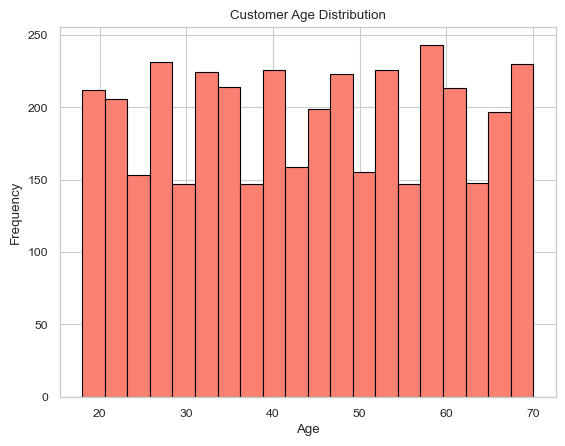
This bar graph showcases how much customers spend on purchases in various locations. It highlights that in places like Arizona and Alaska, customers tend to spend the most on average for their purchases. Meanwhile, Connecticut stands out with the lowest average purchase amount among the locations studied. This information suggests that customers in Arizona and Alaska are more inclined to spend higher amounts compared to those in Connecticut. Understanding these spending patterns across different locations can help businesses tailor their strategies to better serve customers in those areas, ensuring they meet their preferences and needs.



In this bar graph, we see the connection between item categories and the average money spent by customers. The data reveals that customers typically spend around $60 on clothing, footwear, and accessories. Interestingly, outerwear also hovers around the $60 mark, indicating a similar average purchase amount. The graph offers a clear snapshot of how different item categories correlate with the average spending habits of customers, emphasizing the consistent trend around the $60 range for various products.



This bar graph shows how much money customers are spending and how frequently. The majority, around 250 customers, are spending $100. Overall, it appears that the average spending falls in the range of $70 to $100 USD. The graph gives us a quick snapshot of the distribution of purchase amounts, indicating a concentration around the $100 mark. This suggests that a significant portion of customers tends to spend within this range, while fewer customers spend more or less than that amount.



This bar plot showcases the age distribution of customers in the dataset. It's clear that the majority of customers fall within the age range of 20 to 70. Interestingly, those in their 10s exhibit the highest frequency of visits among all age groups. The graph suggests a diverse age representation, with a concentration in the 20 to 70 range. This implies a broad appeal of the products or services to a wide demographic. Understanding this age distribution can guide marketing strategies and product offerings to cater to the varied preferences and needs of customers across different age brackets.



This violin graph showcases how the purchase amount is distributed across various categories. At the $100 mark, clothing emerges as the most popular choice, evident from the widened section of the graph. On the other hand, outerwear finds Favor among buyers at the $20 range, indicating a common price point for such items. The visualization provides insights into customer preferences, revealing that a significant number opt for clothing at higher price points, while outerwear attracts more purchases at a more affordable range around $20. This analysis helps understand the spending patterns and popular price ranges within each category.

**SMART Questions**

**Question - 1:     Which factors (such as customer gender and location) have the most significant impact on the purchase amount?**

**About the T-Test :**

A t-test is a statistical tool for determining if there is a significant difference in the means of two groups. It's especially useful when dealing with tiny sample numbers or an unknown population standard deviation.

T-tests are classified into two types: independent samples t-tests and paired samples t-tests. The independent samples t-test compares the means of two distinct and independent groups, such as test results from two separate courses. The paired samples t-test, on the other hand, evaluates the means of two related groups, similar to before-and-after measurements in an experiment on the same individuals.

The t-test computes a t-value based on the difference in group means and the variability within groups. A higher t-value implies that the difference between groups is less likely to be caused by chance.

The t-value is then compared to a critical value from a t-distribution, taking into account the degrees of freedom and the significance threshold (typically 0.05). If the estimated t-value is more than the crucial value, it indicates that the groups differ significantly.

Overall, t-tests assist researchers and analysts in determining whether reported differences between groups are actual or the product of random fluctuation in the data.

**About Linear Regression:**

Linear regression is a fundamental statistical technique used for modeling the relationship between a dependent variable and one or more independent variables. It assumes that there's a linear relationship between the predictor variables (also called features or independent variables) and the outcome variable (also known as the dependent variable).

The model takes the form of a linear equation:

y = mx + c

Where:

y is the dependent variable (the variable you want to predict).

x is the independent variable (the variable used to make predictions).

m is the slope of the line (how much \(y\) changes when \(x\) changes).

c is the y-intercept the value of y when x is 0.

In multiple linear regression, where there are multiple predictors, the equation becomes:

y = b0 + b1x1 + b2x2 + ………+ bnxn

Where:

y is still the dependent variable.

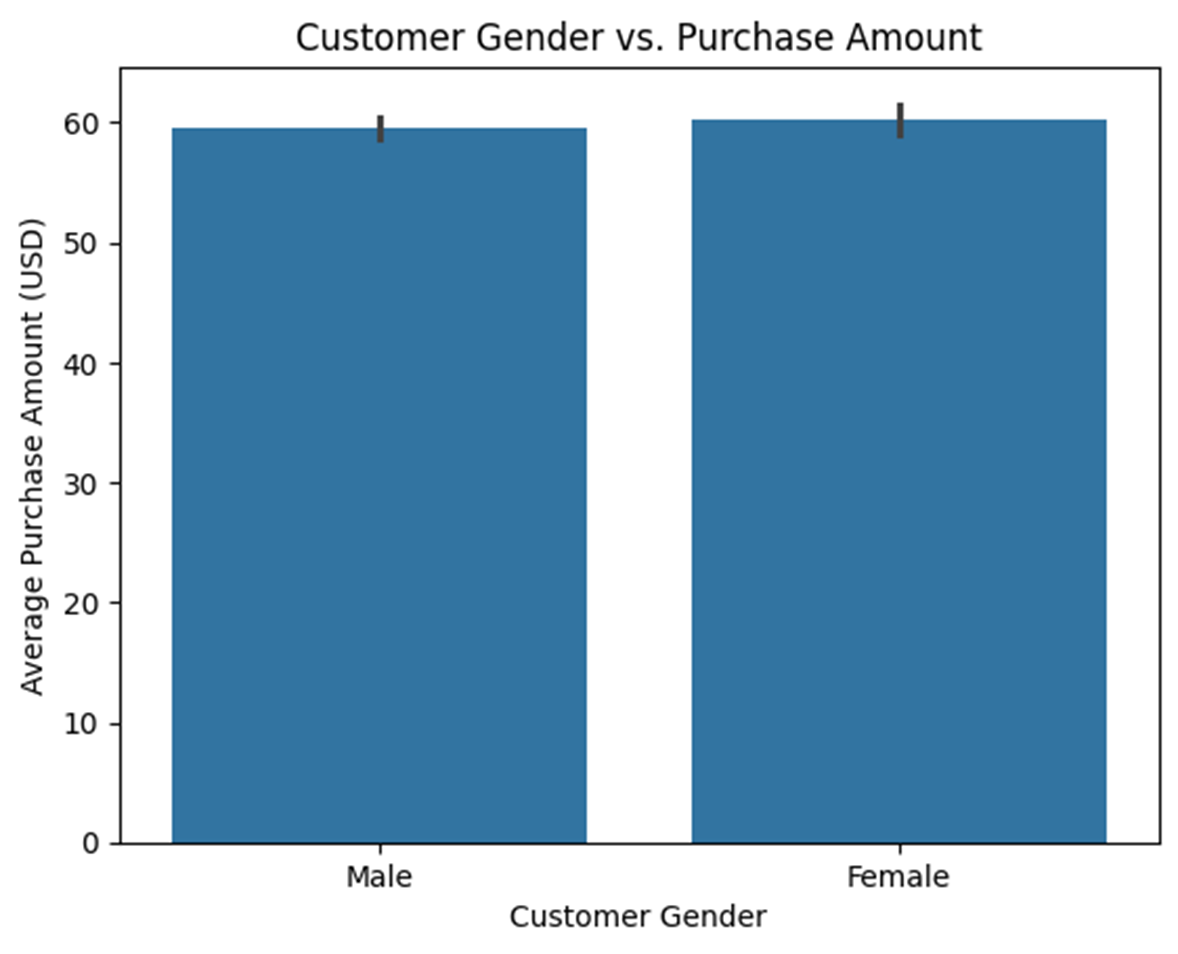
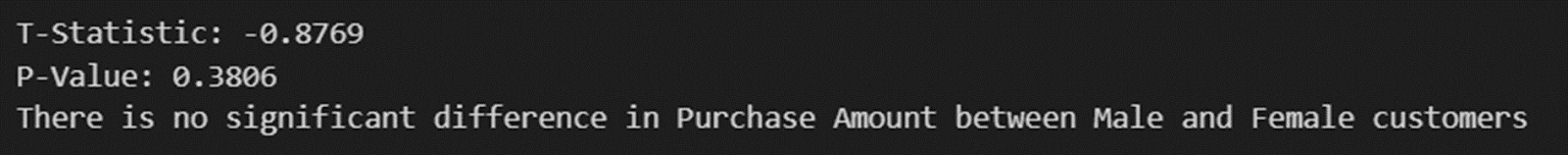
x1, x2, ……. xn are the independent variables.

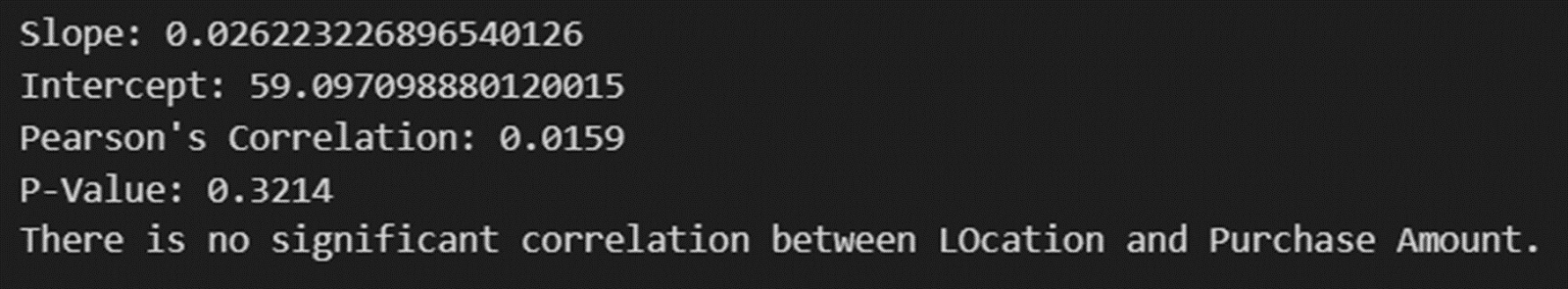
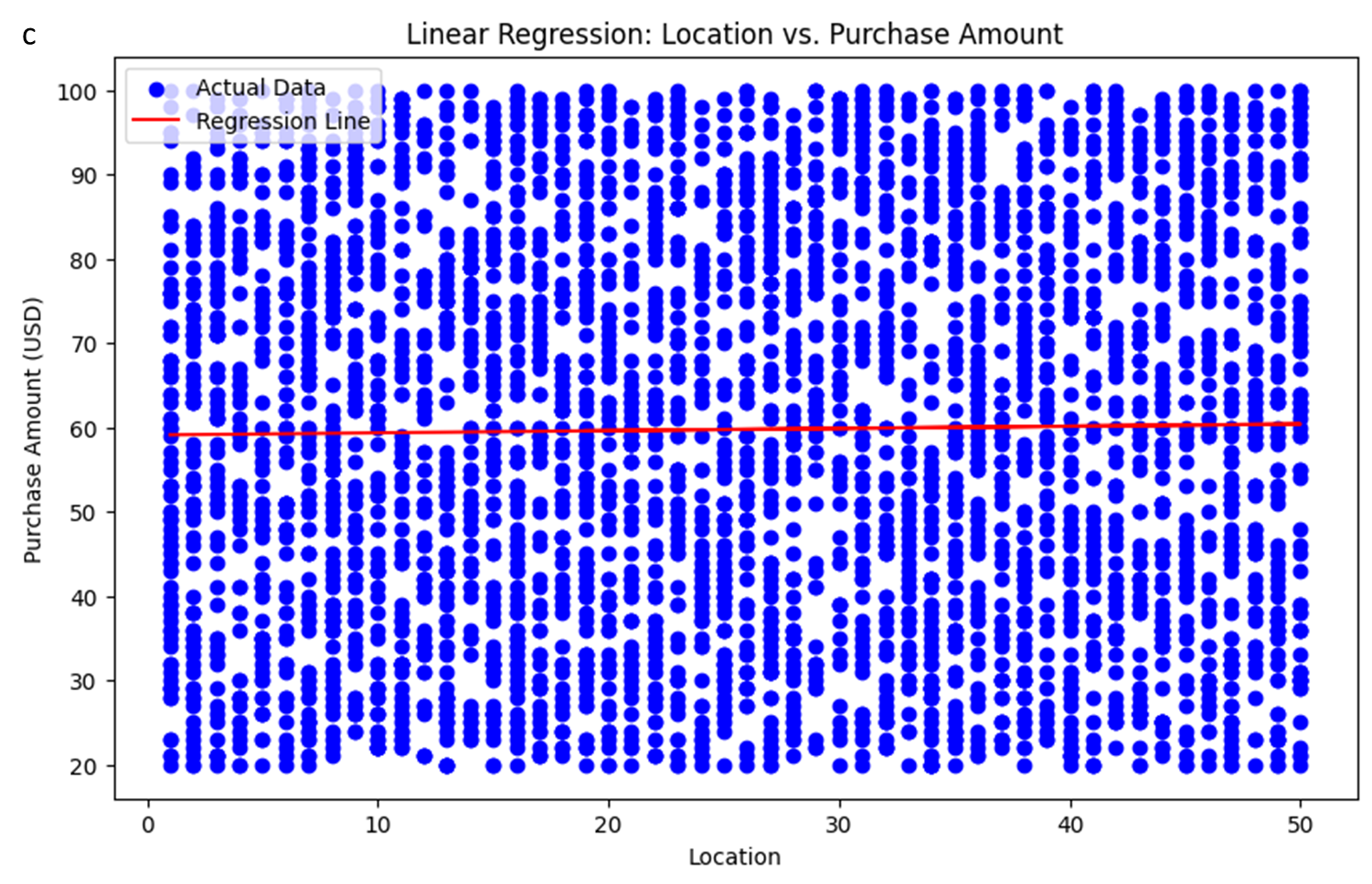
b0 is the intercept.

b1, b2, ……, bn are the coefficients for each independent variable.

The goal of linear regression is to find the best-fitting line (or hyperplane in higher dimensions) that minimizes the difference between the predicted and actual values of the dependent variable. This is often done by minimizing the sum of the squared differences between the observed and predicted values (this method is known as ordinary least squares).

Linear regression is widely used in various fields for predictive analysis, understanding relationships between variables, and making forecasts based on historical data.

****thisgraph explains the relationship between male and female customers with respect to average purcahse amount they spent in a transaction ****

****

**Literature Review**