EXECUTIVE SUMMARY

Analysis on Customer Purchasing behaviour across states in America and product category Performance.
Seasonal trends indicate that customers tend to make more purchases during the winter and autumn months compared to spring and summer, suggesting a potential need for targeted marketing campaigns during these peak seasons.

The Outerwear category exhibits slightly lower purchase amounts compared to other categories, highlighting an area for potential improvement through enhanced product offerings or marketing strategies.

Demographic analysis shows that males account for 67% of total spending, while females contribute 32%, indicating a gender disparity in purchasing behaviour. Additionally, XL sizes have lower purchases compared to sizes like Large, Small, and Medium, suggesting a need to reassess inventory management or product offerings for this size.

The use of promo codes does not significantly impact purchase behaviour, suggesting that other promotional strategies may be more effective.

Geographically, Montana stands out with a remarkable number of customers, indicating a thriving consumer market that should be further capitalized on. Product category analysis reveals that clothing is the most popular across all consumer demographics. Accessories are equally popular across age groups, except for those aged 15-25 and 65-75. Footwear is particularly popular among the 45-55 age group, while outerwear enjoys consistent popularity across all age groups.

Based on these findings, our project proposal includes leveraging at least two separate data sources: sales data from the company's internal database and customer demographic data from market research firms. By integrating these data sources, we can develop targeted marketing strategies, optimize inventory management, and enhance product offerings to maximize profitability and customer satisfaction.

Objective

The Goal for this project is to understand the data set/patterns in order to create more effective targeted marketing campaigns.

Introduction Background

As a n executive marketing manager for a retailer, I was tasked with utilising the data collected from our customers across a number of stores across America to understand the purchasing behaviour in order to create a marketing campaign that is effective.

Method and Process

For this project, I have chosen to solely use Python as a method of analysing the data and producing all the output.
The data is specific to the apparel market covering e.g. footwear, Outwear, Accessories and clothing with a wide range of data points including Subscription, promotion codes, discounts and purchasing frequency. The Goal is to understand customer behaviour using this data. The data set was chosen as it includes all types of data sets i.e. nominal, ordinal, discrete and continuous.

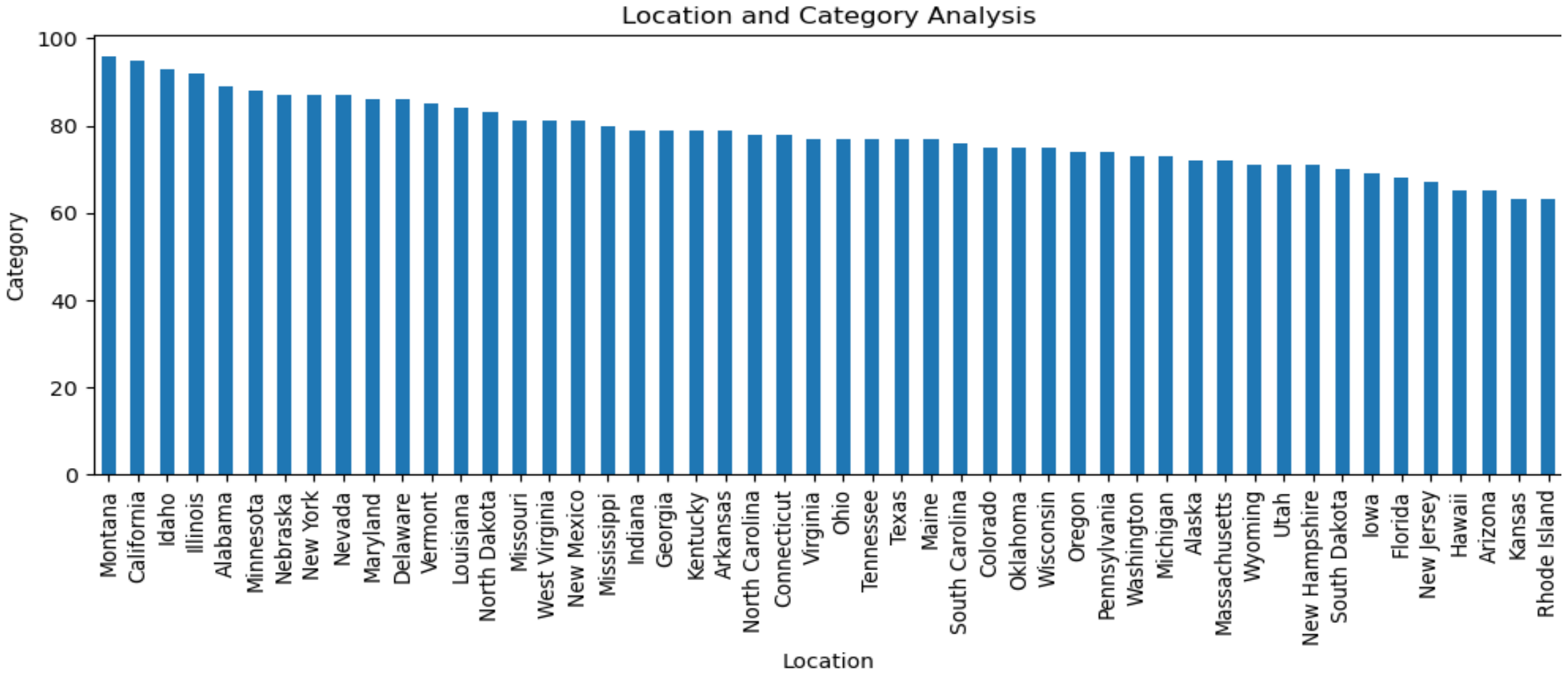
Snapshot of Data Collected (using Python)

Customer ID	3900
Age	53
Gender	2
Item Purchased	25
Category	4
Purchase Amount (USD)	81
Location	50
Size	4
Color	25
Season	4
Review Rating	26
Subscription Status	2
Shipping Type	6
Discount Applied	2
Promo Code Used	2
Previous Purchases	50
Payment Method	6
Frequency of Purchases	7

Age	Gender	Item Purchased	Category	Purchase Amount (USD)	Location	Size	Color	Season	Review Rating	Subscription Status	Shipping Type	Discount Applied	Promo Code Used	Previous Purchases	Payment Method	Frequer Purcha
55	Male	Blouse	Clothing	53	Kentucky	L	Gray	Winter	3.1	Yes	Express	Yes	Yes	14	Venmo	Fortnigl
19	Male	Sweater	Clothing	64	Maine	L	Maroon	Winter	3.1	Yes	Express	Yes	Yes	2	Cash	Fortnigl
50	Male	Jeans	Clothing	73	Massachusetts	S	Maroon	Spring	3.1	Yes	Free Shipping	Yes	Yes	23	Credit Card	Wee
21	Male	Sandals	Footwear	90	Rhode Island	M	Maroon	Spring	3.5	Yes	Next Day Air	Yes	Yes	49	PayPal	Wee
45	Male	Blouse	Clothing	49	Oregon	M	Turquoise	Spring	2.7	Yes	Free Shipping	Yes	Yes	31	PayPal	Annu
46	Male	Sneakers	Footwear	20	Wyoming	M	White	Summer	2.9	Yes	Standard	Yes	Yes	14	Venmo	Wee
63	Male	Shirt	Clothing	85	Montana	M	Gray	Fall	3.2	Yes	Free	Yes	Yes	49	Cash	Quart

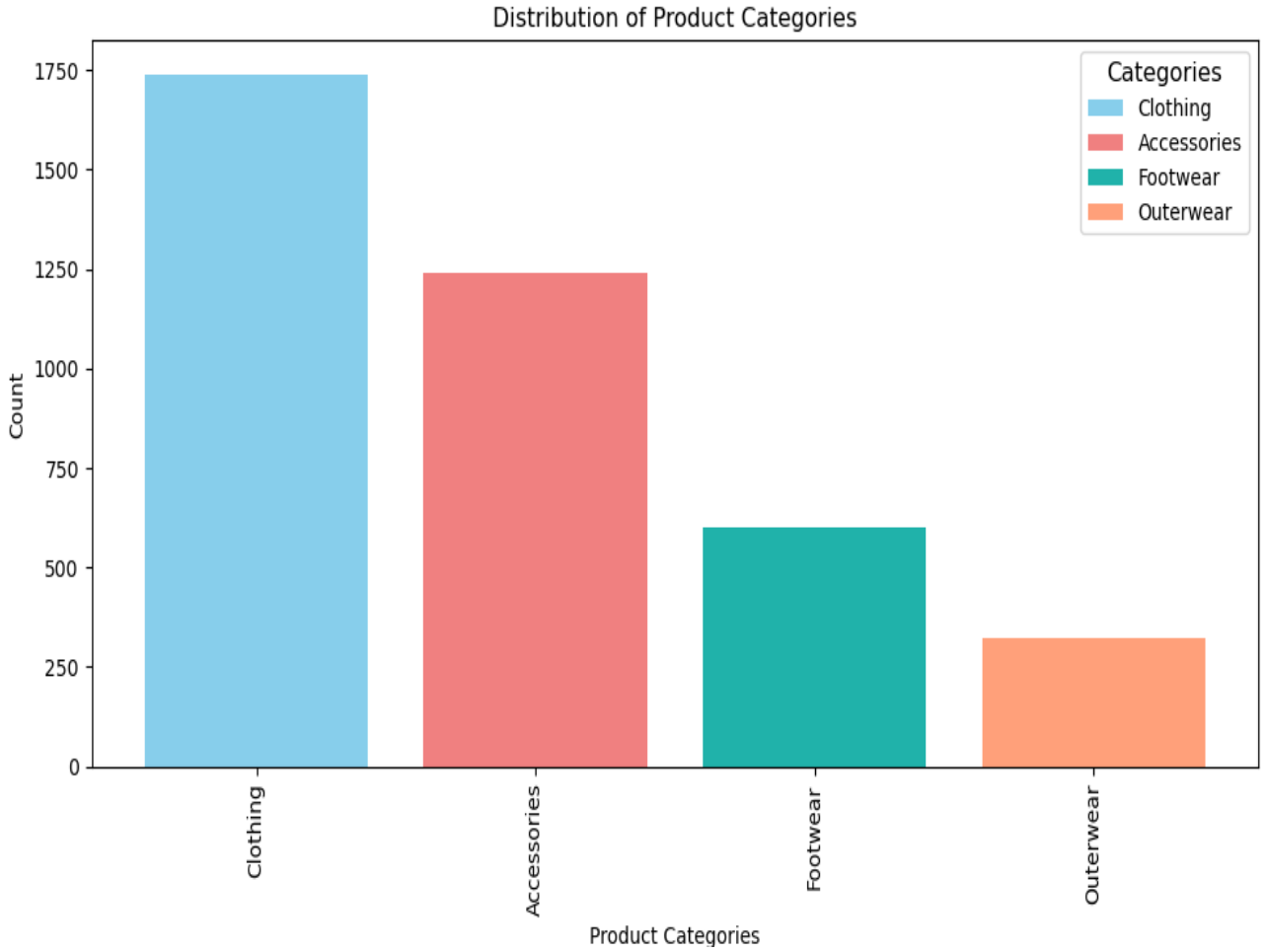
Consumer Behaviour

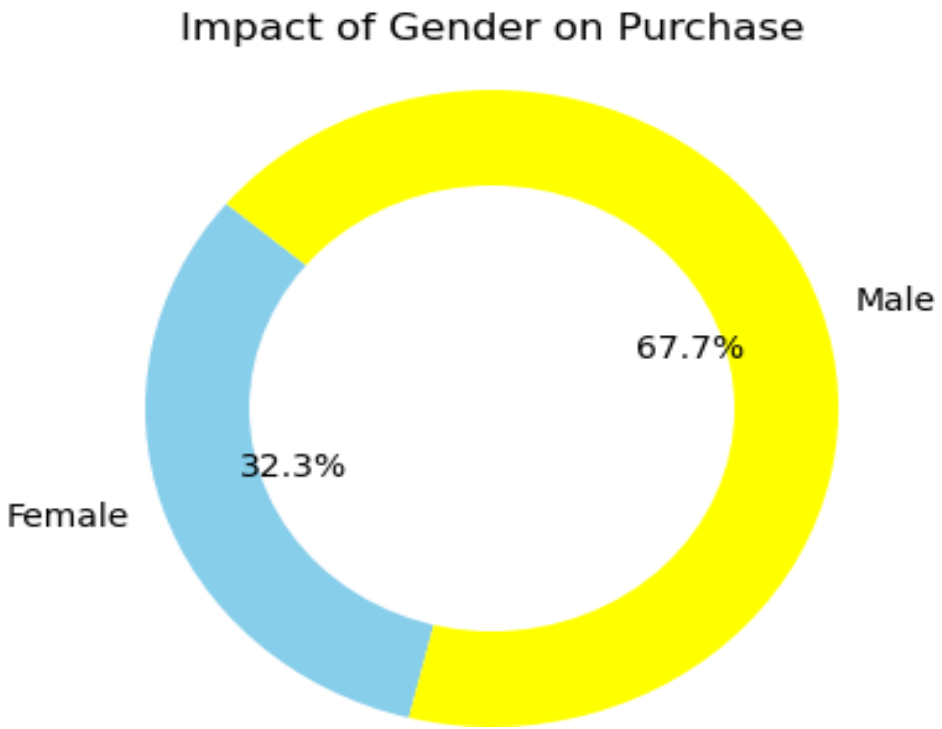
Consumer Analysis across states in America showing Montana stands out with a remarkable number of customers, indicating a thriving consumer market.



Category and Gender Analysis

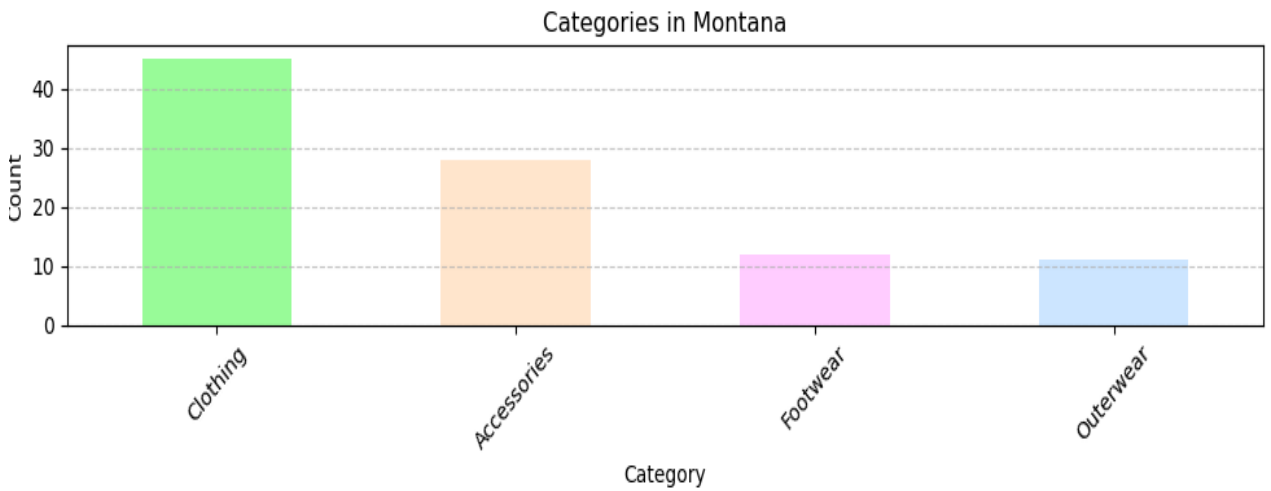
- Clothing is the most popular product category across all consumer demographics.
- Males account for 67% of the total spending, while females contribute 32% to purchases.

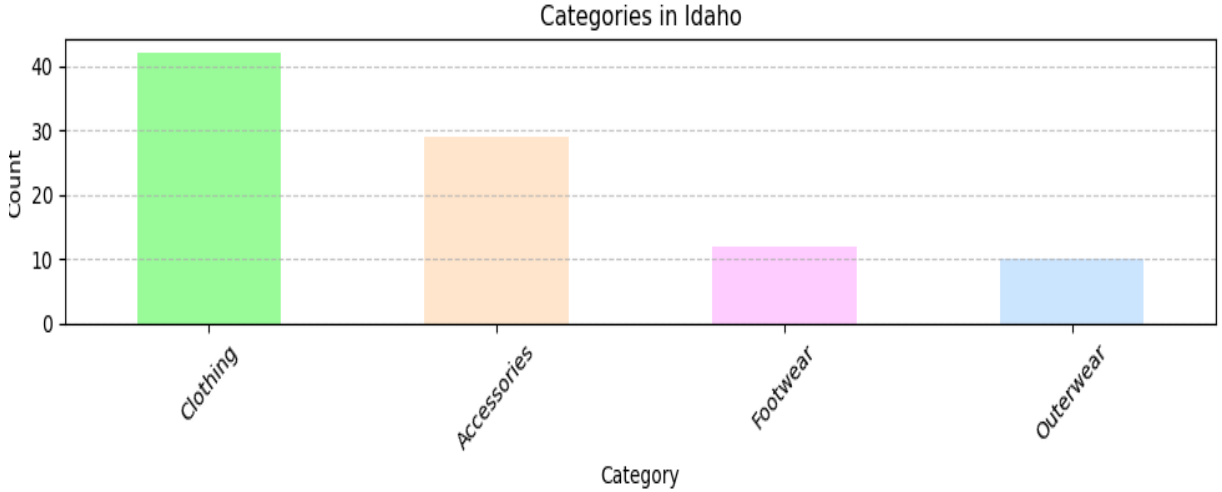


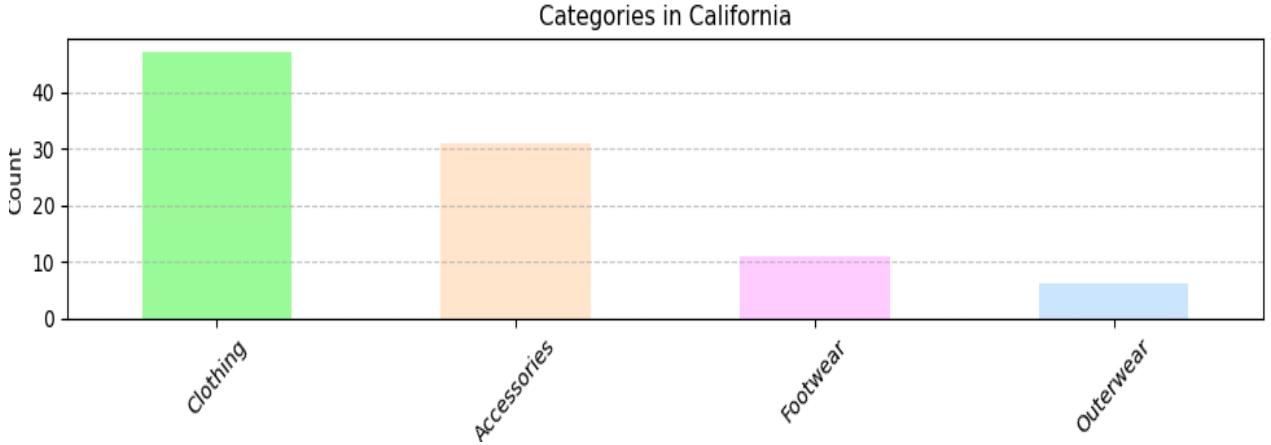


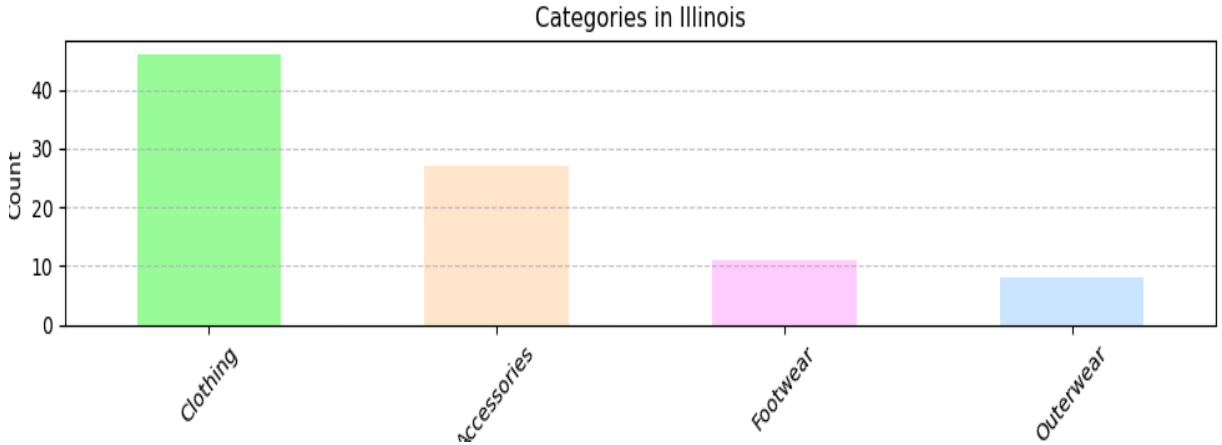
Tops States Category Analysis

- Outerwear enjoys consistent popularity across all age groups.
- The Outerwear category has slightly lower purchase amounts compared to other categories, indicating potential areas for improvement.









Analysis Methodology

- Below are examples of the Python Code used in the analysis.
- In analysing categories across each state, this was done by creating subplots for each location and then counting the number of categories for each location. To make it presentable, I made sure to create different colour schemes as shown further above.
- When analysing regional trends, this was done by grouping categories by each Location
- GitHub Location: https://github.com/Isaacserungoui/American-States-Customer-Behavior-Data-Analysis

Distribution of Product Category Code

```
category_counts = df['Category'].value_counts()

# Define a list of different colors for each bar
colors = ['skyblue', 'lightcoral', 'lightseagreen', 'lightsalmon', 'lightpink']

# Create a figure and axis
plt.figure(figsize=(10, 6))
ax = plt.gca()

# Plot the bar chart with different colors for each bar
bars = plt.bar(category_counts.index, category_counts.values, color=colors)

# Add labels and title
plt.xlabel('Product Categories')
plt.ylabel('Count')
plt.title('Distribution of Product Categories')
plt.xticks(rotation=90)

# Display the chart
plt.tight_layout()

# Optionally, you can add a legend to show the correspondence between colors and categories
legend_labels = category_counts.index[:len(colors)]  # Take labels for the number of colors used
legend = plt.legend(bars[:len(colors)], legend_labels, title='Categories', loc='upper right')
plt.setp(legend.get_title(), fontsize=12)

plt.show()
```

Subplot Grid per Location Code

```
top_locations = df['Location'].value_counts().head(5).index

# Define different colors for bars
colors = ['#98FB98', '#FFE5CC', '#FFCCFF', '#CCE5FF', '#9467bd', '#8c564b', '#e377c2', '#7f7f7f', '#bcbd22', '#17becf']

# Create a subplot grid for each location
fig, axes = plt.subplots(5, 1, figsize=(10, 15))

# Iterate through the top locations and create category distribution plots with different colors
for i, location in enumerate(top_locations):
    location_data = df[df['Location'] == location]

    # Count the most common product categories in this location
    category_counts = location_data['Category'].value_counts().head(10)

    # Create a bar plot for the category distribution with different colors
    ax = axes[i]
    category_counts.plot(kind='bar', ax=ax, color=colors)
    ax.set_title(f"Categories in {location}")
    ax.set_xlabel("Category")
    ax.set_ylabel("Count")
    ax.set_xticklabels(category_counts.index, rotation=45)
    ax.grid(axis='y', linestyle='--', alpha=0.7)

# Adjust subplot layout for a clean appearance
plt.tight_layout()

# Display the visualizations
plt.show()
```

Further Analysis Methodology

- It was important to get a Summary of each location by a few data points including Average Purchase amount, Most popular category in addition to the online preferences.
- This was done by grouping each location and the calculating the average purchase amount in addition to the popular categories in each regions as shown below.

```
location_groups = df.groupby("Location")

# Analyze regional trends
for location, location_data in location_groups:
    print(f"Regional Trends for {location}:")

    # Calculate average purchase amount in this region
    avg_purchase_amount = location_data["Purchase Amount (USD)"].mean()
    print(f"Average Purchase Amount: ${avg_purchase_amount:.2f}")

    # Count the most popular product categories in this region
    popular_categories = location_data["Category"].value_counts().idxmax()
    print(f"Most Popular Category: {popular_categories}")

    # Analyze online shopping preferences
    online_shopping = location_data["Shipping Type"].apply(lambda x: "Online" if "Express" in x or "Standard" in x else "Offline")
    online_percentage = (online_shopping.value_counts() / len(online_shopping)) * 100
    print(f"Online Shopping Preference:")
    print(online_percentage)

    # Consider other factors based on your data and business context
    print("\n")
```

```
Regional Trends for Alabama:
Average Purchase Amount: $59.11
Most Popular Category: Clothing
Online Shopping Preference:
Shipping Type
Offline     68.539326
Online      31.460674
Name: count, dtype: float64


Regional Trends for Alaska:
Average Purchase Amount: $67.60
Most Popular Category: Clothing
Online Shopping Preference:
Shipping Type
Offline     58.333333
Online      41.666667
```

Summary of Results & Conclusion

- Montana stands out with a remarkable number of customers, indicating a thriving consumer market.
- Clothing is the most popular product category across all consumer demographics.
- Accessories are equally popular across age groups, except for those aged 15-25 and 65-75.
- Footwear is particularly popular among the 45-55 age group.
- Outerwear enjoys consistent popularity across all age groups.
- Males account for 67% of the total spending, while females contribute 32% to purchases.
- XL size exhibits lower purchases compared to other sizes like Large, Small, and Medium.
- The usage of promo codes doesn't seem to have a significant impact on purchase behaviour.
- Customers tend to make more purchases during winter and autumn compared to spring and summer.
- The Outerwear category has slightly lower purchase amounts compared to other categories, indicating potential areas for improvement.

THANK YOU