# My work

I have uploaded the following files for you to use: index.csv I uploaded the file: index.csv complete this project

# run\_code

# Importing necessary libraries

import

pandas

as

pd

# Loading the CSV file into a DataFrame

file\_path

=

'index.csv'

df

=

pd.read\_csv(file\_path,

encoding='ascii')

# Displaying the head of the DataFrame to understand its structure

print(df.head())

The dataset contains transaction records from a coffee shop, capturing details such as the date, time, payment method, card identifier, transaction amount, and the type of coffee purchased.

# Data Overview

* **Columns (Features):**
* date: The date of the transaction (YYYY-MM-DD)
* datetime: The exact timestamp of the transaction
* cash\_type: Payment method, either 'cash' or 'card'
* card: Card identifier or None if paid by cash
* money: Transaction amount in monetary units
* coffee\_name: Name of the coffee purchased

# Content Summary

* The dataset records transactions at a coffee shop over a period spanning from late March 2024 to late July 2024.
* Transactions include both cash and card payments.
* Card transactions involve anonymized card IDs (e.g., 'ANON-XXXX-XXXXXXXX').
* The money values range approximately from $18.12 to $40.00.
* Multiple entries per day, with some customers making multiple purchases.
* The coffee\_name includes various coffee types such as Latte, Cappuccino, Espresso, Americano, and their variants with Milk.

Now, let's perform some basic analysis to understand the total sales and the most popular coffee types. I'll calculate the total sales and the count of each coffee type sold.

# run\_code

# Calculating total sales and the count of each coffee type sold

# Total sales

total\_sales

=

df['money'].sum()

# Count of each coffee type

coffee\_counts

=

df['coffee\_name'].value\_counts()

# Displaying the results

print('Total Sales: $'

+

str(total\_sales))

print(coffee\_counts)

# run\_code

# Now, let's visualize the total sales and the count of each coffee type sold u import matplotlib.pyplot as plt

# Setting up the figure plt.figure(figsize=(12, 6))

# Bar chart for coffee counts

plt.bar(coffee\_counts.index, coffee\_counts.values, color='#1B47A7')

# Adding titles and labels plt.title('Coffee Sales Count', pad=15) plt.xlabel('Coffee Type', labelpad=10) plt.ylabel('Count Sold', labelpad=10) plt.xticks(rotation=45)

# Displaying the plot

plt.grid(color='#6D6D6D', linestyle='--', linewidth=0.5) plt.gca().set\_facecolor('#1A1A1A') plt.gcf().set\_facecolor('#1A1A1A') plt.show()

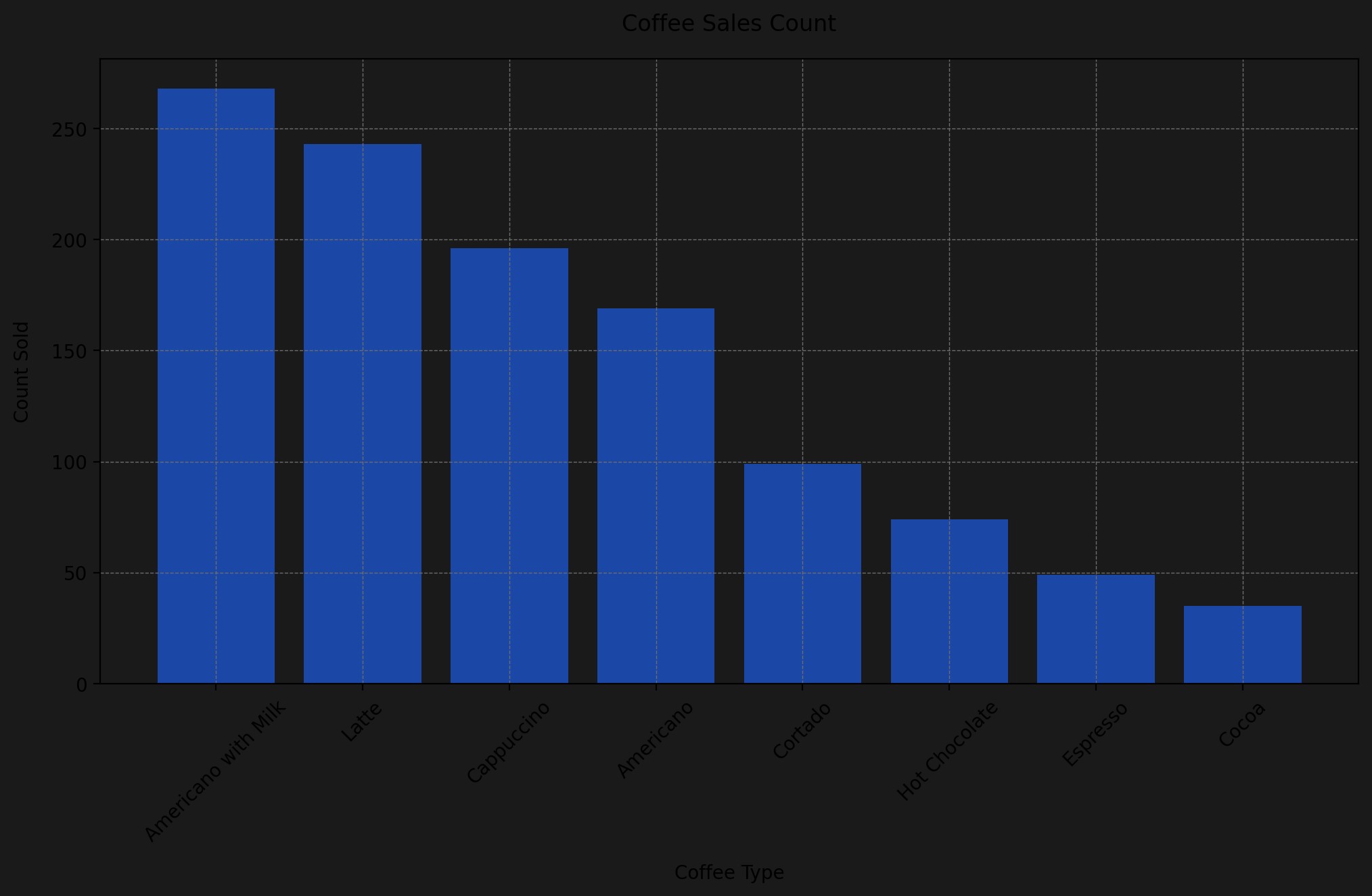
The total sales from the coffee shop transactions amount to approximately $37,508.88. Additionally, I've provided a count of each coffee type sold, which highlights the most popular items among customers.

Here are the counts of coffee sold:

|  |  |
| --- | --- |
|  | **count** |
| Americano with Milk | 268 |
| Latte | 243 |
| Cappuccino | 196 |
| Americano | 169 |
| Cortado | 99 |
|  | **count** |
| Hot Chocolate | 74 |
| Espresso | 49 |
| Cocoa | 35 |

Next, I created a bar chart to visualize the sales count of each coffee type. This visualization helps to quickly identify which coffee types are the most popular among customers.

Here is the bar chart displaying the coffee sales count:



From the chart, we can see that "Americano with Milk" is the most sold coffee type, followed closely by "Latte" and "Cappuccino." This information can be useful for inventory management and marketing strategies.

If you need further analysis or specific insights from this dataset, please let me know!