

```
import pandas as pd

# Load the sales data
sales_data = pd.read_csv('sales_data.csv')

# Display the first few rows of the dataframe
sales_data.head()
```

```
↗
```

	Date	Time	StoreID	CustomerID	OrderID	Product Name	Size	Crust
0	2024-12-05	02:56:32	0	0	0	BBQ Chicken	medium	thick
1	2024-12-05	02:56:32	0	0	0	Margherita	medium	thin
2	2024-12-05	02:56:32	0	0	1	Margherita	medium	thin
3	2024-12-05	02:56:32	0	0	1	Hawaiian	large	stuffed crust
4	2024-12-05	02:56:32	0	0	1	Supreme	large	thin

```
# Count the occurrences of each product
product_counts = sales_data['Product Name'].value_counts()

# Display the most prevalent products
print(product_counts.head())
```

```
↗ Product Name
Veggie      1176
Pepperoni   1143
Margherita  1136
Meat Lovers 1124
Hawaiian    1117
Name: count, dtype: int64
```

```
# Define a large basket as having more than a certain number of items
large_basket_threshold = 5
```

```
# Group by CustomerID and count the number of items in each basket
basket_sizes = sales_data.groupby('CustomerID').size()
```

```
# Count the number of large baskets
large_basket_counts = basket_sizes[basket_sizes > large_basket_threshold].count()
```

```
# Display the frequency of large buyers
print(large_basket_counts)
```

```
↗ 10
```

```
# Group by StoreID and count the number of large baskets in each store
large_basket_stores = sales_data[sales_data['CustomerID'].isin(basket_sizes[basket_sizes > large_basket_threshold].index)]
store_large_basket_counts = large_basket_stores['StoreID'].value_counts()
```

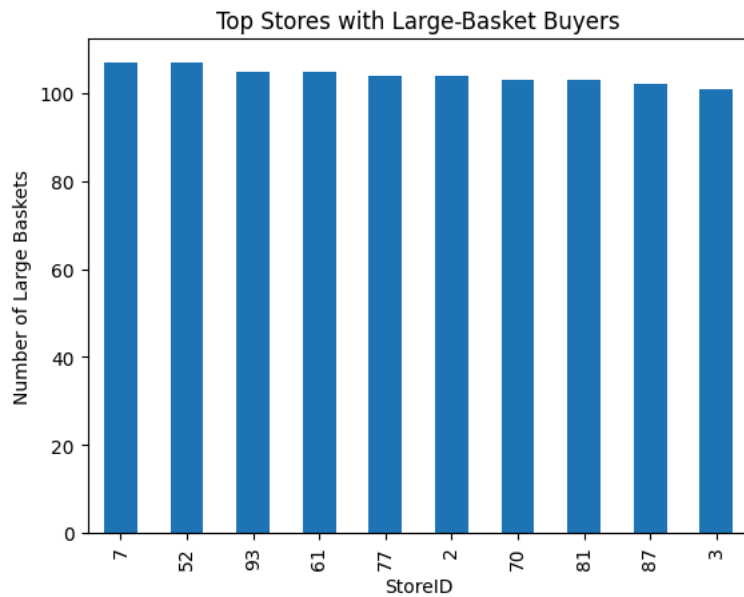
```
# Display the stores with large-basket buyers
print(store_large_basket_counts)
```

```
↗ StoreID
7      107
52     107
93     105
61     105
77     104
...
44     75
65     75
8       75
76     74
20     70
Name: count, Length: 100, dtype: int64
```

```
import matplotlib.pyplot as plt
```

```
# Plot the top stores with large-basket buyers
store_large_basket_counts.head(10).plot(kind='bar')
plt.title('Top Stores with Large-Basket Buyers')
plt.xlabel('StoreID')
```

```
plt.ylabel('Number of Large Baskets')
plt.show()
```



```
# Get the products in large baskets
large_basket_products = large_basket_stores['Product Name'].value_counts()
```

```
# Display the top-N products
top_n = 10
print(large_basket_products.head(top_n))
```



```
Product Name
Veggie      1176
Pepperoni   1143
Margherita  1136
Meat Lovers 1124
Hawaiian    1117
BBQ Chicken 1107
Supreme     1097
Cheese      1097
Name: count, dtype: int64
```

```
# Group by CustomerID and get the average basket makeup
basket_makeup = sales_data.groupby('CustomerID')['Product Name'].apply(lambda x: x.value_counts(normalize=True))
```

```
# Display the average categorical makeup of baskets
print(basket_makeup.head())
```



```
CustomerID
0      Pepperoni    0.148649
      Veggie       0.137387
      Cheese       0.131757
      Hawaiian    0.130631
      Meat Lovers  0.120495
Name: Product Name, dtype: float64
```

```
# Plot the categorical makeup of baskets
basket_makeup_df = basket_makeup.unstack().mean().sort_values(ascending=False)
basket_makeup_df.plot(kind='bar')
plt.title('Average Categorical Makeup of Baskets')
plt.xlabel('Product Name')
plt.ylabel('Average Proportion in Basket')
plt.show()
```

