



# Phase 4

# Market Basket insights

Team members:

Nivya shaji  
Aleena Suresh  
Abira S  
Anushya L

# What is a CMS market basket?

Although “market basket” technically describes the mix of goods and services used in providing health care, this term is also commonly used to denote the input price index (that is, cost category weights and price proxies combined) derived from that market basket. Accordingly, the term “market basket” as used in this document refers to the various CMS input price indexes. A CMS market basket is described as a fixed-weight, Laspeyres-type index because it measures the change in price, over time, of the same mix of goods and services purchased in the base period.

Which items are frequently purchased together by customers?



### Shopping Baskets



Customer 1



Customer 2



Customer 3



Customer *n*

Market Analyst

# Types of Market Basket Analysis

Market basket analysis is classified into two types:

Predictive Market Basket Analysis

Differential Market Basket Analysis

# Benefits of Market Basket Analysis



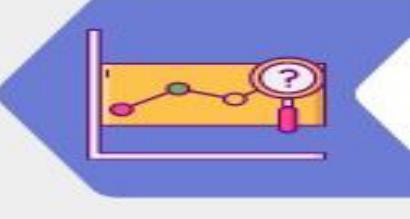
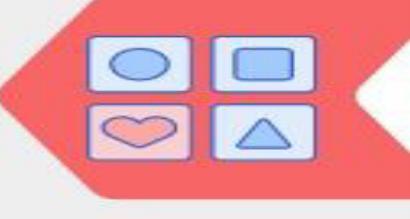
# Predictive Market Basket Analysis

Predictive market basket analysis is a type of data mining technique that uses historical data on customer purchases to make predictions about future customer behavior. The goal of predictive market basket analysis is to identify items that are likely to be purchased together and use this information to inform business decisions such as product placement, marketing strategies, and inventory management.

# Differential Market Basket Analysis

Differential Market Basket Analysis (DMBA) is a statistical technique used to identify the difference between two or more market baskets, or sets of items, typically purchased together by customers. It is commonly used in retail and marketing to understand the purchasing behavior of customers, as well as to identify trends and patterns in sales data. The goal of DMBA is to find items that are unique to each market basket and determine the associations between them, which can then be used to inform promotional strategies, product placement, and other marketing decisions.

# Market Basket Analysis Types

1.  **Descriptive market basket analysis**
2.  **Predictive market basket analysis**
3.  **Differential market basket analysis**

# Applications of Market Basket Analysis

## The Retail Industry

Market basket research can assist retailers to find goods that are commonly purchased together, which can help them make product placement, marketing, and price decision

## Financial Services and Banking

Market basket analysis can be used by banks and financial organizations to evaluate client data and uncover trends in their purchasing habits.

```
import pandas as pd
import numpy as np

# Load the transaction data from a CSV file (you can replace 'transactions.csv' with your file)
Data = pd.read_csv('transactions.csv')

# Display the first few rows to inspect the data
Print(data.head())

# Preprocessing steps:
# 1. Handle missing values (if any)
Data = data.dropna() # Drop rows with missing values

# 2. Convert date columns to datetime objects (if applicable)
Data['date'] = pd.to_datetime(data['date'])

# 3. Perform any necessary data transformations or calculations
Data['total_amount'] = data['quantity'] * data['unit_price']

# Now, your data is loaded and preprocessed.

# You can perform various analysis or visualization tasks on this preprocessed data.
```

## 1. Load Data in Pandas

To work on the data, you can either load the CSV in Excel or in Pandas. For the purposes of this tutorial, we'll load the CSV data in Pandas.

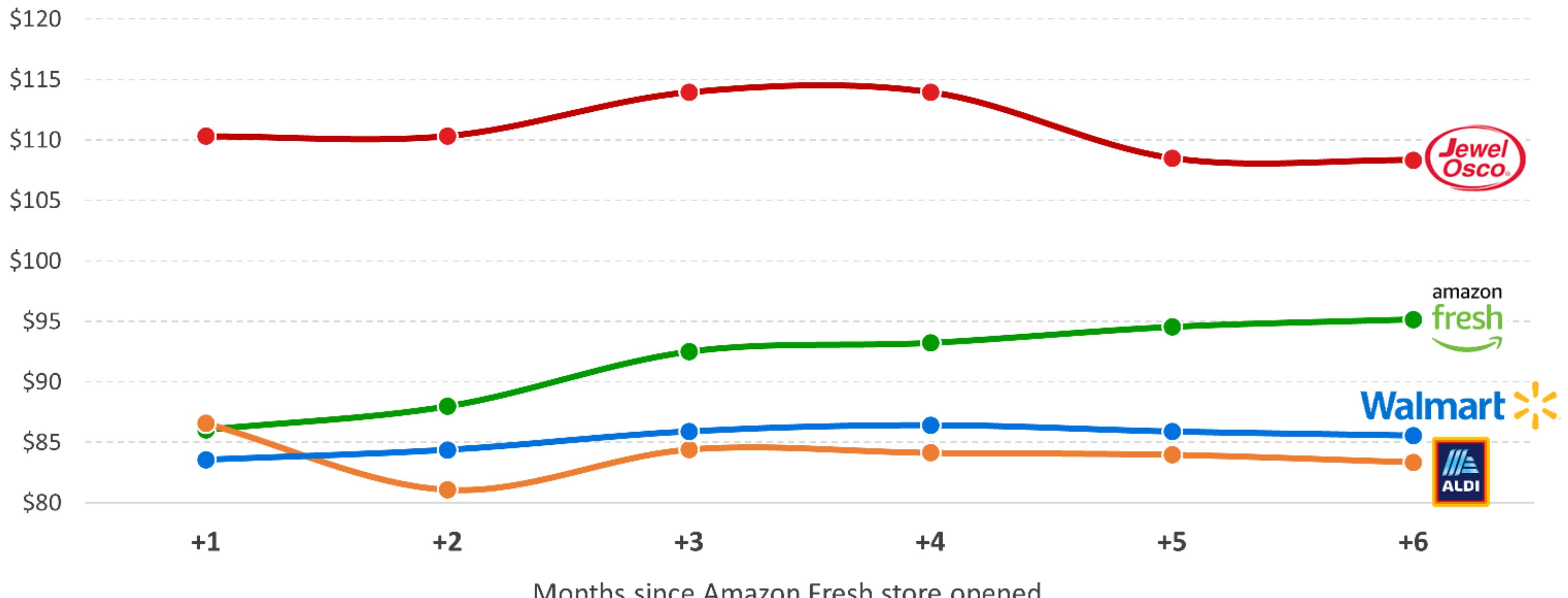
```
Df = pd.read_csv('train.csv')
```

Let's take a look at the data format below:

```
>>> df.info()
<class 'pandas.core.frame.DataFrame'>
Int64Index: 891 entries, 0 to 890
Data columns (total 12 columns):
PassengerId 891 non-null int64
Survived 891 non-null int64
Pclass 891 non-null int64
Name 891 non-null object
Sex 891 non-null object
Age 714 non-null float64
SibSp 891 non-null int64
Parch 891 non-null int64
Ticket 891 non-null object
Fare 891 non-null float64
Cabin 204 non-null object
Embarked 889 non-null object
```

If you carefully observe the above summary of Pandas, there are 891 total rows but Age shows only 714 (which means we're missing some data), Embarked is missing two rows and Cabin is missing a lot as well. Object data types are non-numeric so we have to find a way to encode them to numerical values.

**Fig A. Customer Basket Cost at Select Grocers Over Six-Month Period Since Amazon Fresh Opened**





# Thank you