

Anticipating Customer Purchases

Bentley Ou
December 2019

Think About...



What if you know exactly what product a customer will purchase...

And...

Their purchase can be made with just one click away!

Market Basket Analysis

Market Basket Analysis



What

Business Intelligence

A key technique used to uncover relationships between products using association learning

Why

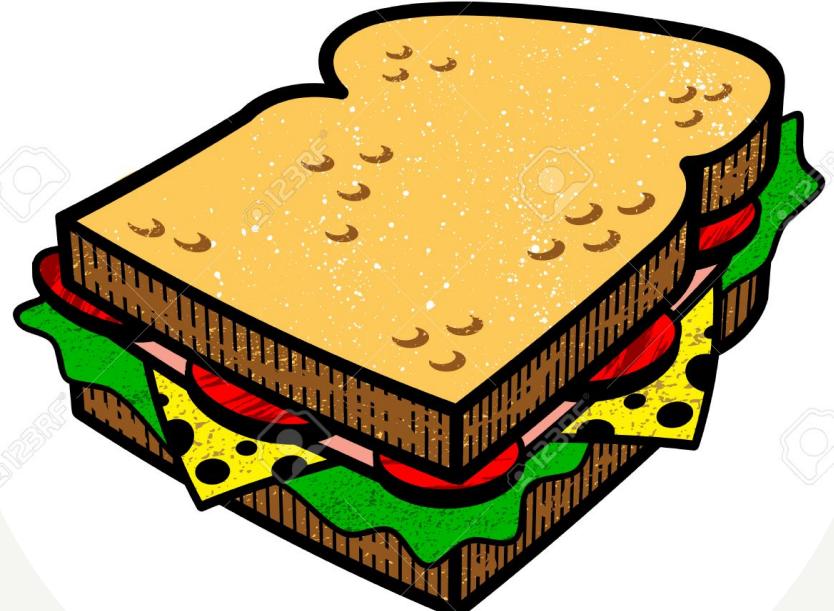
Business Strategy

1. Cross Sell
2. Recommendations
3. Marketing Promotions

So What?

Insights to help better understand - and ultimately serve - customers by discovering purchasing patterns

For Example



Sandwich

A customer first buys a sandwich



Cookie

And buys a cookie



Beverage

Then buys a drink on the
CONDITIONAL that they bought a
sandwich AND cookie

Market Basket Analysis



How it works

Support

Frequency

How many times were the combination of a sandwich, a cookie, and a beverage bought altogether

Confidence

Conditional

How many times were the purchase of a beverage made given the purchase of a sandwich and a cookie

Lift

Relationship or Chance

Ratio of how many purchases of a sandwich, a cookie, and a drink were actually made to the expected number of purchases as if each product were independent

InstaCart

Flow



Exploratory Data Analysis

1. AWS to support 32+ million rows
2. Jupyter Notebook
3. Matplotlib and Seaborn



Build Association Rules

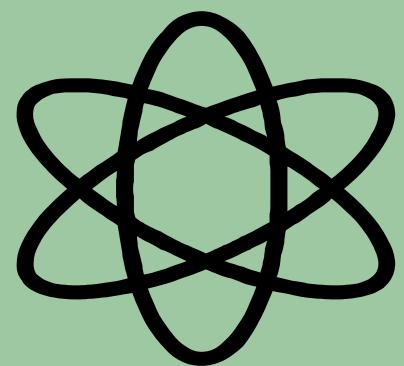
1. Python generators
2. Support, confidence, and lift
3. Association Learning



Application Deployment

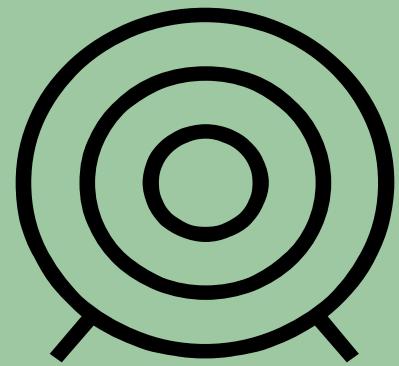
1. NLP for product lookup
2. HTML, CSS, and Python
3. Heroku Hosting

Key Findings



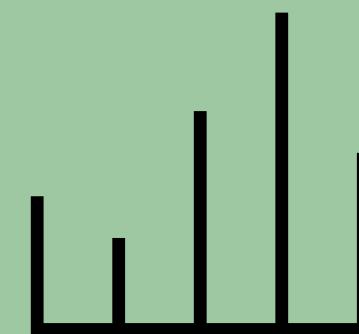
Department

30% of all purchases are from the **produce** department



Reorder

More than **80%** of milk and dairy products are repurchases



Purchasing Period

Customers tend to make purchases after the **7th, 14th, or 21st** day from a prior purchase



Association Mining

1

ITEM SET CREATION

Create item sets from all customer product transaction history

2

FILTER

Set support threshold at 0.01

3

METRIC

Use Lift > 1 and high confidence to determine highly associated products

Web Application

Flask Application on Heroku



PRODUCT LOOKUP ALGORITHM

1. CountVectorizer
2. Cosine Similarity
3. Threshold > 0.95



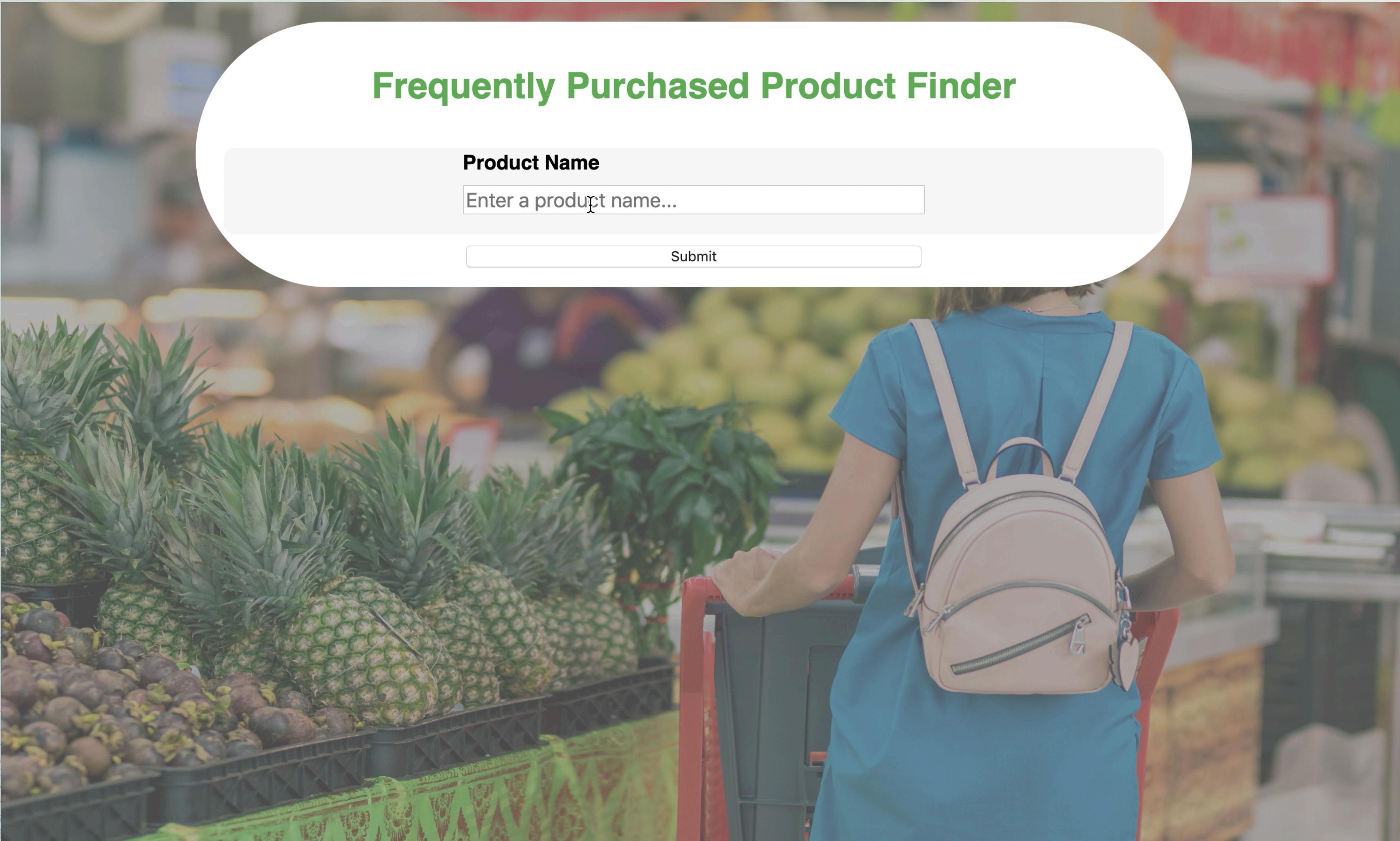
FIND ASSOCIATIONS

1. Used association rules
2. Filtered on Confidence and Lift
3. Returns product recommendations



FLASK APP

1. Deployed on Heroku



Other Use Cases

Market Basket Analysis

Industrial

Manufacturing

Predictive analysis of equipment failure

Biology

Medicine

Discovery of co-occurrence relationships among diagnosis and pharmaceutical active ingredients prescribed to different patient groups

Finance

Risk

Fraud detection based on credit card usage data



A large, faint watermark-like graphic in the bottom left corner consists of numerous thin, light gray lines forming a complex pattern of overlapping triangles and abstract shapes.

Thank you Questions?

Bentley Ou
December 2019