

**ibm.biz/
localcart**



Analyse

Customer Data

and Recommend

Products with

PixieDust

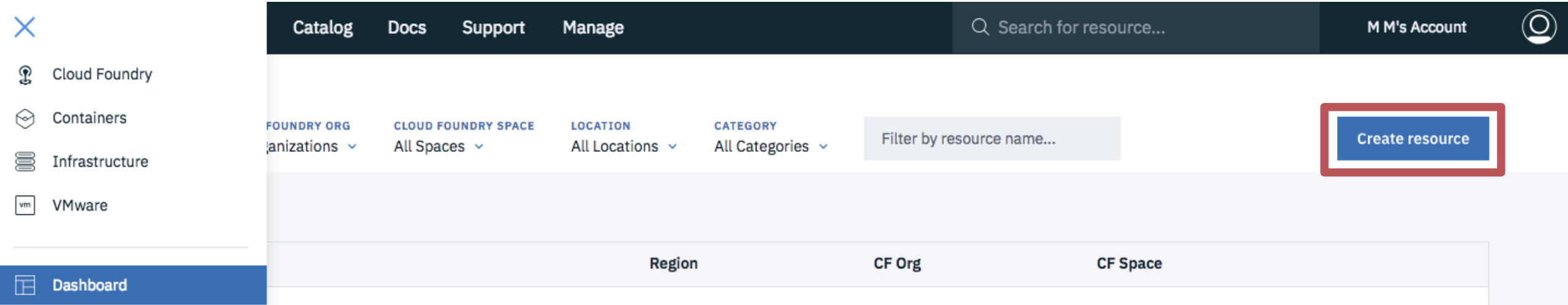
IBM Developer

Arlemi Turpault Yamini Rao Margriet Groenendijk Mofe Salami



Access Watson Studio

At the top of your IBM Cloud dashboard click Create Resource



Search for Watson Studio and click on the tile

Select the Lite plan and click Create

The screenshot shows the IBM Cloud Catalog interface. At the top, there's a dark navigation bar with 'IBM Cloud', 'Catalog', 'Docs', 'Support', and 'Manage'. A search bar on the right contains 'Search for resource...'. On the far right, there's a user profile icon and 'M M's Account'. Below the navigation bar, a search bar contains 'label:lite watson' and a 'Filter' button. On the left, a sidebar lists categories: 'All Categories (17)', 'Compute', 'Containers', 'Networking', 'Storage', 'AI (13)', 'Analytics (1)', 'Databases', 'Developer Tools', 'Integration', and 'Internet of Things (1)'. The main area displays three service tiles: 'Text to Speech', 'Tone Analyzer', and 'Visual Recognition', all with 'Lite • IBM' plans. Below these, the 'Watson Studio' tile is highlighted with a red border. It also shows the 'Lite • IBM' plan and a description: 'Embed AI and machine learning into your business. Create custom models using your own data.'

IBM Cloud Catalog Search Results for **label:lite watson**

Watson Studio
Lite • IBM

Embed AI and machine learning into your business.
Create custom models using your own data.

Go back to the dashboard and click on your Watson Studio service and then click Get Started

Services

Name ▾	Location	Resource Group	Plan	Details	Service Offering	
Watson Studio-ng	United Kingdom	Default	Lite	Provisioned	Watson Studio	⋮
cloud-object-storage-tz	global	Default	Lite	Provisioned	Cloud Object Stor...	⋮
pm-20-sr	United Kingdom	Default	Lite	Provisioned	Machine Learning	⋮

Part 0: Getting Started

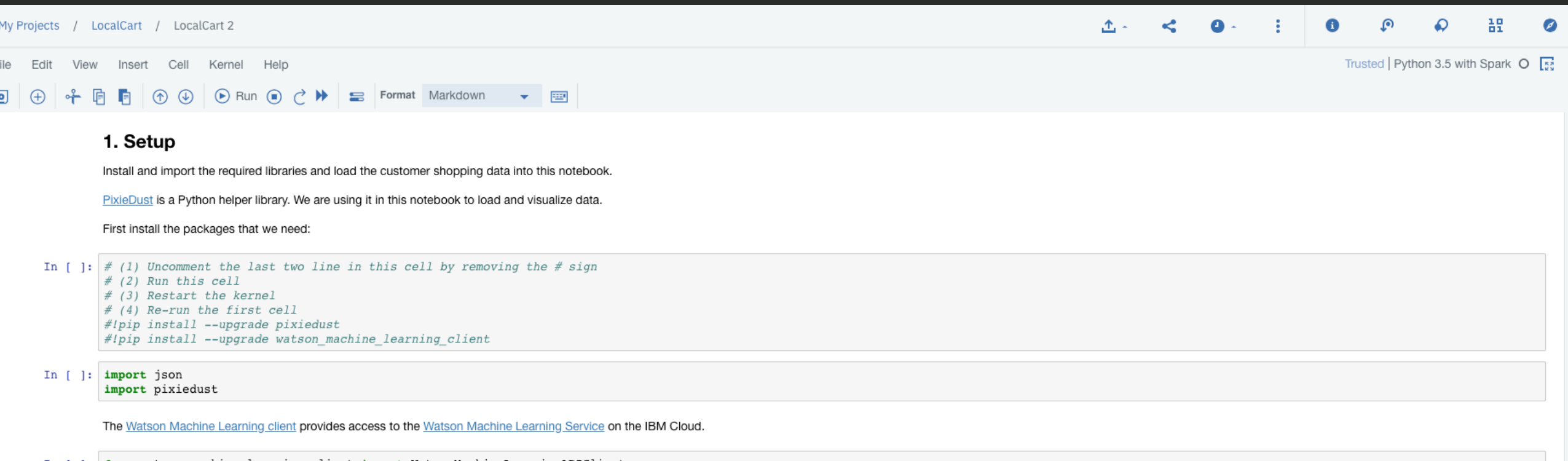
**ibm.biz/
localcart**

Part 1: Static data analysis using Python, Apache Spark, and PixieDust

IBM Developer

Jupyter Notebooks - <http://jupyter.org/>

- Notes
- Ideas
- Python code
- Charts and tables



The screenshot shows a Jupyter Notebook interface. The top bar includes navigation links for 'My Projects', 'LocalCart', and 'LocalCart 2'. The left sidebar contains menu items: 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', and 'Help'. The right sidebar shows 'Trusted | Python 3.5 with Spark' and a 'Format' dropdown menu set to 'Markdown'. The main content area is titled '1. Setup' and contains the following text:

Install and import the required libraries and load the customer shopping data into this notebook.

[PixieDust](#) is a Python helper library. We are using it in this notebook to load and visualize data.

First install the packages that we need:

```
In [ ]: # (1) Uncomment the last two line in this cell by removing the # sign
# (2) Run this cell
# (3) Restart the kernel
# (4) Re-run the first cell
#!pip install --upgrade pixiedust
#!pip install --upgrade watson_machine_learning_client
```

```
In [ ]: import json
import pixiedust
```

The [Watson Machine Learning client](#) provides access to the [Watson Machine Learning Service](#) on the IBM Cloud.

Why Spark?

Cluster computing platform designed to be **fast** and **general** purpose

APIs in Python, Java, Scala and SQL

Data Science: data wrangling, machine learning

Data Processing: data pipelines

Why Spark?

```
df = df.select()
```

```
df.filter()
```

```
df.join()
```

```
df.groupby().sum()
```

```
df.printSchema()
```

```
df.show()
```

```
df.cache()
```


Why Spark?

```
df = df.select()
```

```
df.filter()
```

```
df.join()
```

```
df.groupby().sum()
```

```
df.printSchema()
```

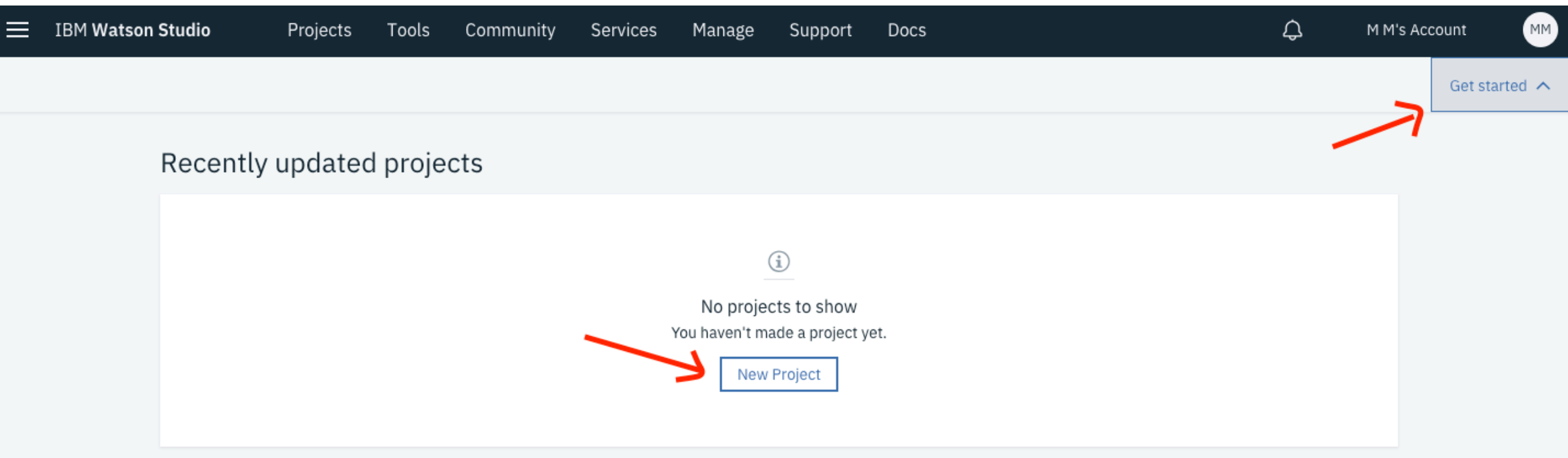
```
df.show()
```

```
df.cache()
```

Lazy!

```
import pixiedust  
display(df)
```

Watson Studio - create a new project



Watson Studio - create a notebook

New notebook

Blank

From file

1From URL

Name*

2LocalCart 1

39 Characters Remaining

Description

Type your Description here

Notebook URL*

3https://github.com/IBMCodeLondon/localcart-workshop/blob/master/notebooks/part-1-analyze-ci

4

Services

spark-be

Environments

✓

Default Spark Python 3.5 XS (Driver with 1 vCPU and 4 GB RAM, 2 executors with 1 vCPU and 4 GB RAM each)

Default Spark Scala 2.11 (Driver with 1 vCPU and 4 GB RAM, 2 executors with 1 vCPU and 4 GB RAM each)

Default Spark R 3.4 (Driver with 1 vCPU and 4 GB RAM, 2 executors with 1 vCPU and 4 GB RAM each)

Default Python 3.5 Free (1 vCPU and 4 GB RAM)

Default Python 3.5 XS (2 vCPU and 8 GB RAM)

Default R 3.4 XS (2 vCPU and 8 GB RAM)

Default Python 3.5 S (4 vCPU and 16 GB RAM)

Default R 3.4 S (4 vCPU and 16 GB RAM)

5

Cancel

Create Notebook

Part 1: Static data analysis using Python, Apache Spark, and PixieDust

**ibm.biz/
localcart**

Part 2: Build a product recommendation engine

IBM Developer

Watson Machine Learning

Last meetup: train model in Watson Studio



IBM Watson Studio

Projects

Tools

Catalog

Community

Services

Manage

Support

Docs

Notifications

Margriet Groenendijk

1687107 - Margriet Groenendijk's Account



New model

Define model details

Name

Model name

100

Description

Model description

300

Machine Learning Service

pm-20-jq



Select model type

☒ Model builder ☐ From file ☐ From sample

Select runtime

Only Spark environments supporting Scala kernels can be used for model builder creation.

Select runtime from the list



Automatic

Prepare my data and create a model automatically

Manual

Let me prepare my data and select which models to train

Need something more flexible? Create a [notebook](#) or design a [Modeler flow](#)

Watson Machine Learning

Today: train model in a notebook

Customers who viewed this item also viewed these products



Dualit Food XL1500
Processor

\$560

 Add to cart



Kenwood kMix Manual
Espresso Machine



\$250

 Select options



Weber One Touch Gold
Premium Charcoal
Grill-57cm

\$225

 Add to cart



NoMU Salt Pepper and
Spice Grinders

\$3

 View options

Spark MLlib

- **ML Algorithms**: common learning algorithms such as classification, regression, clustering, and collaborative filtering
- **Featurization**: feature extraction, transformation, dimensionality reduction, and selection
- **Pipelines**: tools for constructing, evaluating, and tuning ML Pipelines
- **Persistence**: saving and load algorithms, models, and Pipelines
- **Utilities**: linear algebra, statistics, data handling, etc.

<https://spark.apache.org/docs/2.2.2/ml-guide.html>

Recommender Engine

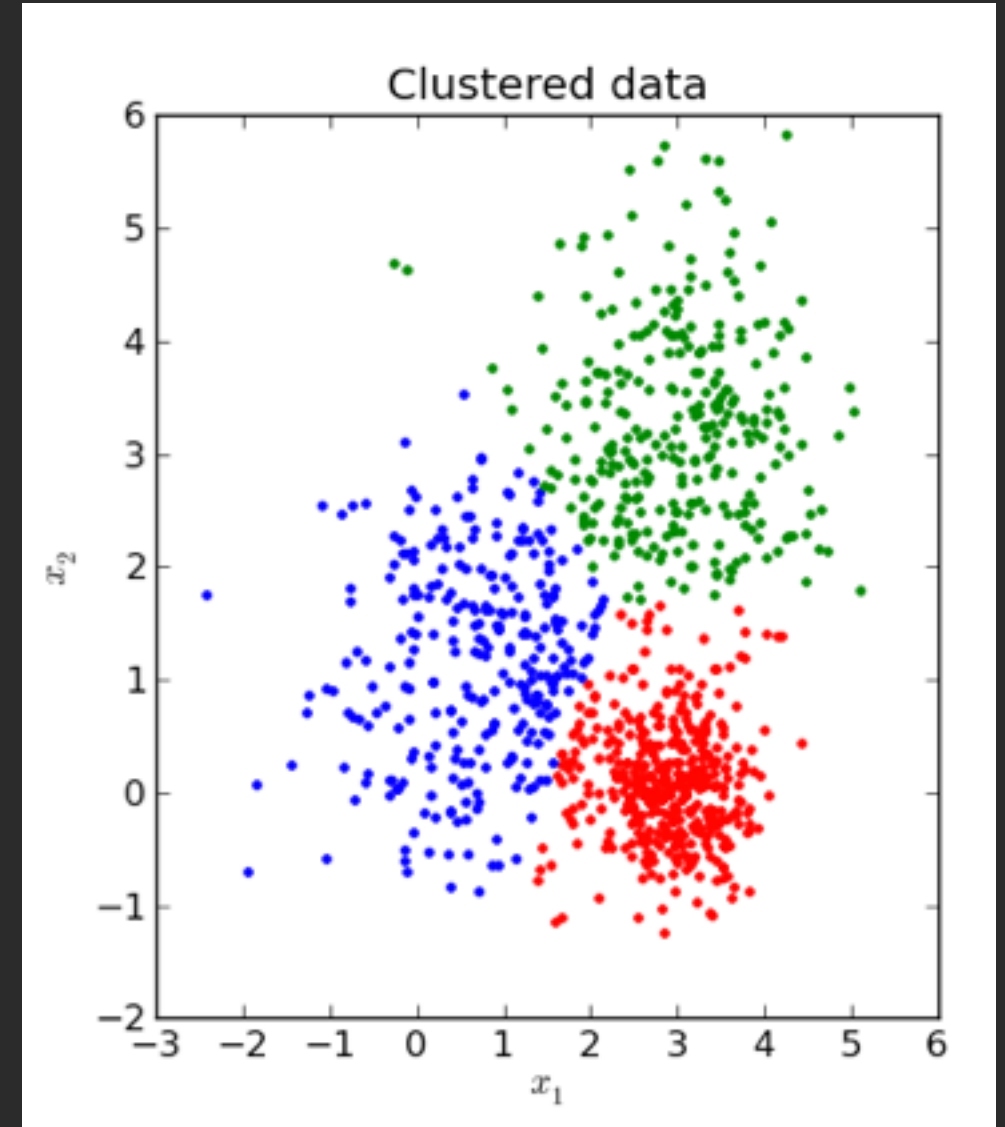
Every dot is a customer

The axes are the number of products A and B bought

You can now use clustering algorithms to group these customers

Imagine not 2 but thousands of products in a thousands dimensional space

With machine learning algorithms you can still find clusters of customers who are similar



But how do you get from the clusters to a recommendation?

Customers who viewed this item also viewed these products



Dualit Food XL1500
Processor

\$560

 Add to cart



Kenwood kMix Manual
Espresso Machine




\$250

 Select options



Weber One Touch Gold
Premium Charcoal
Grill-57cm

\$225

 Add to cart



NoMU Salt Pepper and
Spice Grinders

\$3

 View options

But how do you get from the clusters to a recommendation?

Recommend the most popular items in the cluster


But filter out the products already bought

Customers who viewed this item also viewed these products



Dualit Food XL1500
Processor

\$560

 Add to cart



Kenwood kMix Manual
Espresso Machine



\$250

 Select options



Weber One Touch Gold
Premium Charcoal
Grill-57cm

\$225

 Add to cart



NoMU Salt Pepper and
Spice Grinders

\$3

 View options

Watson Machine Learning

- Spark MLlib
- scikit-learn
- XGBoost
- TensorFlow
- Keras
- Caffe
- PyTorch

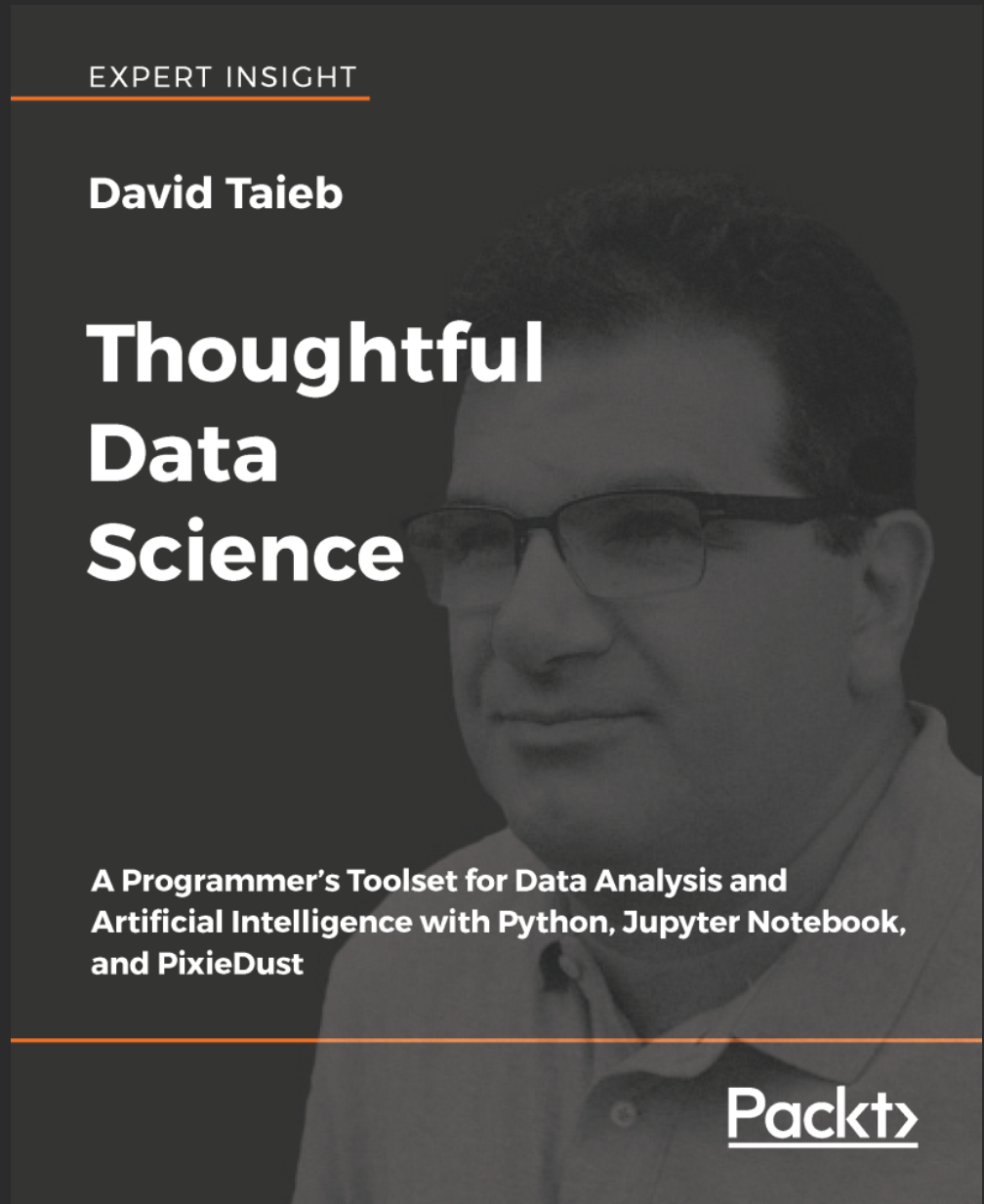


PixieApps

Part 2: Build a product recommendation engine

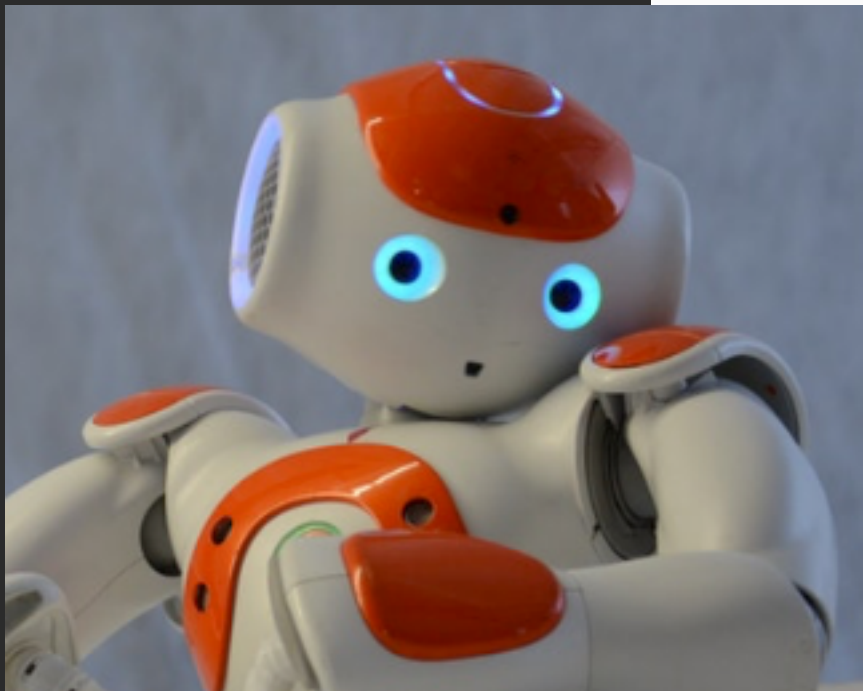
**ibm.biz/
localcart**

PixieDust and PixieApps book!



Build your own AI applications

27 November



Ed Shee

Serverless
Workshop

5 December

