

Cohort 31 - Group 4

Vanessa Gleason, Alistair Marsden, Olivier Rochaix, Eduard Stalmakov

Capstone Project: **Exploratory Questions**

Ideas

Agile sprint

Demographics - census - vanessa find something that applies

Music data set - million song dataset -olivier see if you can get it in a workable form

lastFM - eduard - see what it is like

research in music recommendations

Kafka - new songs added

Dashboard - someone listening to music , based on what they are listening to, you would like to listen

What is the best machine learning model for this?

Who uses music streaming services?

Who doesnt? Could we get them?

Age - what type of music

Age - do they stream music?

What kind of metrics can we get from a persons music listening preferences?

Assuming a customer listened to music for a period of time, (10 songs a day for a year) what is a way to show their favorite song? For the year, week

We can mix a data flow with at machine learning that pulls in from the data flow

For customer A, what is the statistics for this - 10 top songs

What's the most popular song (in our ML)

On the dashboard, have this is the most popular genre, etc.

given a seed playlist title and/or initial set of tracks in a playlist, can we predict the subsequent tracks in that playlist?

How does ecommerce affect economies nationally/by state?

How does e-commerce differ by category (in US) sales revenue ecommerce vs total (food, electronics, sporting goods, etc)

Hypothesis /answer shipping - are people buying more if they are further from stores? Want more choices? Why do people choose ecommerce?

Who purchases on ecommerce? geographic/zipcodes, gender, race, economic status

Which retailers use ecommerce? Which retailers are most successful on ecommerce?

When do consumers ecommerce? Ex. What is Amazons most popular day of the week for sales? Why? Or time or day or time of year?

Machine learning - can we predict if shoppers will purchase a product?

<https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset>

Where are ecommerce store /distribution centers located? Are they clustered somewhere(s)?

Apply questions to specific category - like technology and/or a specific retailer -like amazon

1. *Start broad w this is ecommerce then narrow down to this is a specific example*
2. *Be specific enough that it is interesting*
3. *Recommender - could be unsupervised LM - goes into Marketing/google analytics, will need sales data, click data possibly make a dashboard for individuals buy and click history to predict what they would buy*
4. *Do a supervised LM - easier to incorporate census data, bring in the FDIC data.*
 - a. *Possible questions*
5. *E Commerce impact on economy*
 - a. *# of employees in ecommerce*
 - b. *Effect of in-person retail establishments*
 - i.
 - ii. *People with which characteristics are more likely to purchase via ecommerce?*
 - iii. *What categories of products are more popular with specific demographic types?*

- iv. *Are there demographics groups that are emerging ecommerce shoppers?*
- v. *Overall, what are the most popular type of item(s) people purchase via ecommerce?*
- vi. *Ecommerce over time - are people buying more/less who is buying*