social computing

UNIVERSITÄT DUISBURG ESSEN

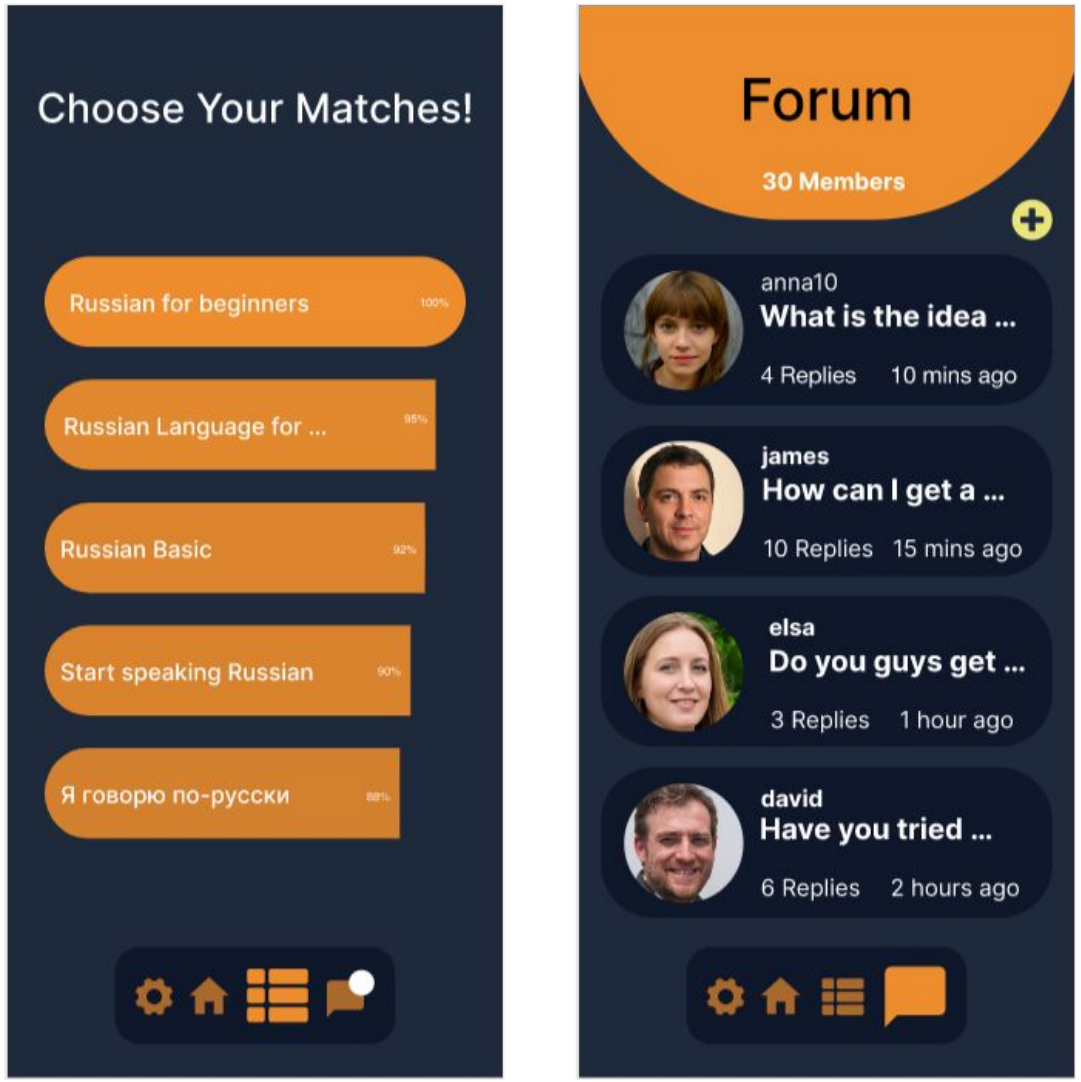
Open-Minded

Presented by Cassandra Spark ✨
Nurmalita Manggali, Nave Wibowo, Jean Qussa, Ali Amini

Overview

- What is SMATCH?
- What - Why - How
- Conceptual Solution
- Implementation
- Result (Demo)
- Summary

What is SMATCH?



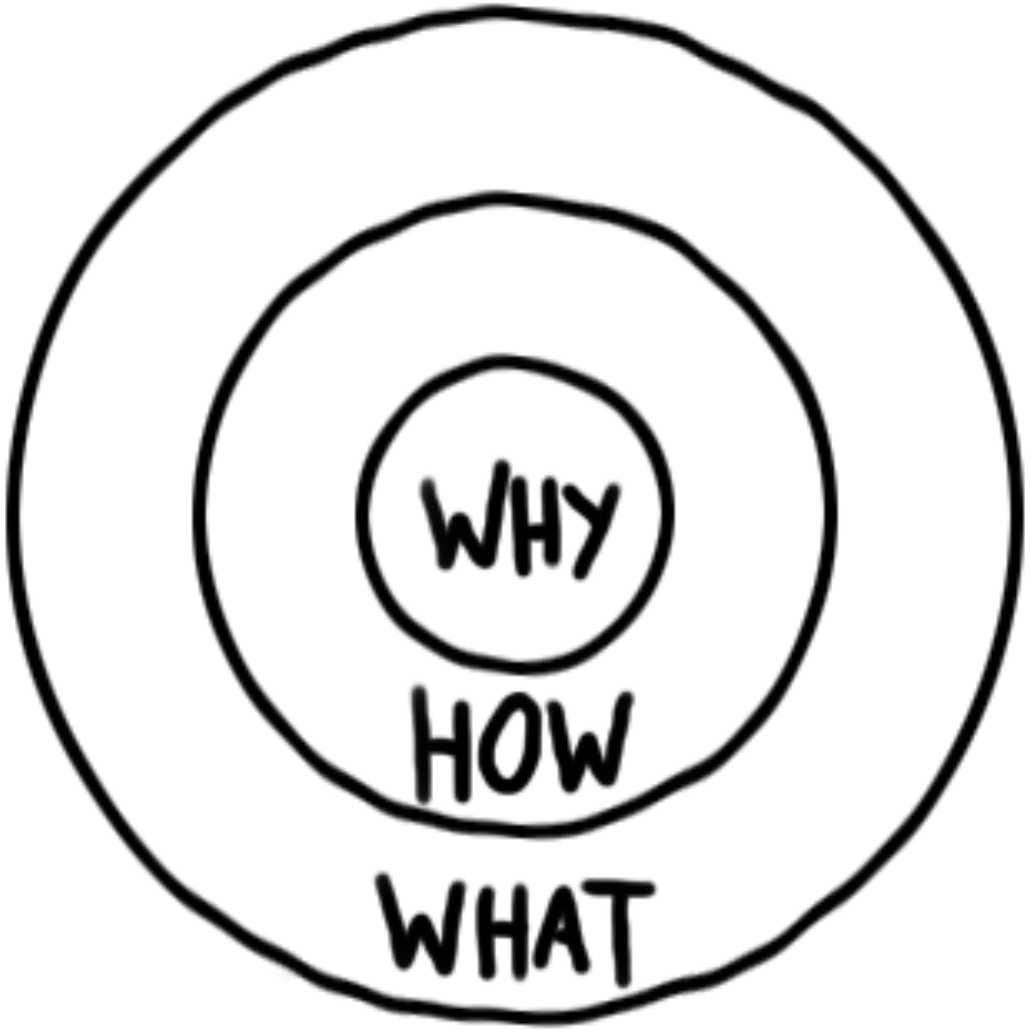
What if we have a matchmaking app that instead of looking for a date, it matches us with appropriate course to take in a MOOC platform

Abstract task description:

An application for students to

- Give a list of suitable classes to take based on preferences
- Connect with other learners in a forum
- Show overviews of multiple platforms in one page

What - Why - How

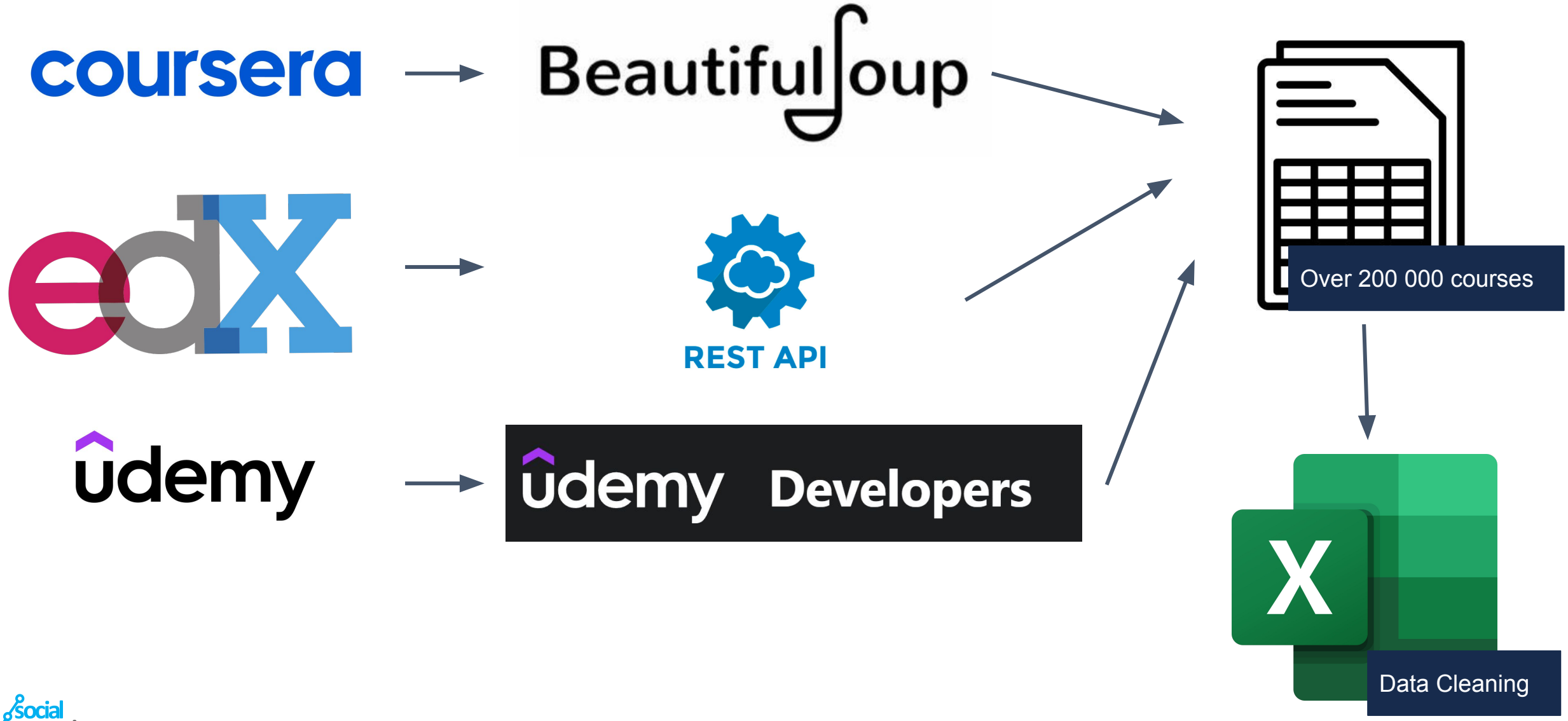


WHAT DATA DO WE HAVE

- Details of courses scraped from various MOOCs platforms
  - Categorical : categories, provider
  - Ordered : price, duration (quantitative), level (ordinal)

- Dataset type : tables, trees
- dataset availability: Static

Data source


coursera
BeautifulSoup
edX
REST API
udemy
udemy Developers
Over 200 000 courses
Data Cleaning

What - **Why** - How

WHY

- is there a way to aggregate courses from different platforms and compare them to ease the choosing process?

Action - Target pairs

- **Discover features** of various courses based on interests
- **Locate distribution** of courses based on interest
- **Compare similarities** and trend between different platforms
- **Summarise trends** of courses between different platforms
- **Derive** the perfect courses from the list for the user

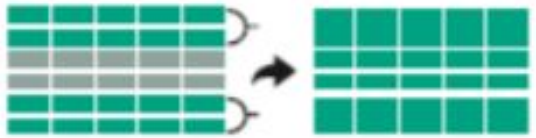
How to present the data

-

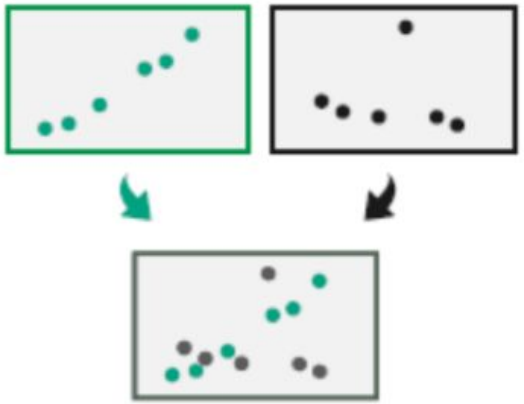
Filter

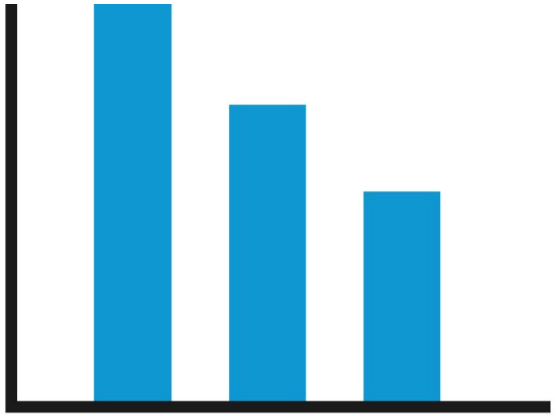
Select

Aggregate



Superimpose





Bar chart

Design Choice

What: data	Categorical, ordered (quantitative)
Why: task	Discover features, locate distribution, Compare similarities, summarise trends, derive
How: Facet	**Superimpose**
How: Reduce	**Filter, Aggregate**
How: Manipulation	**Select, Navigate**

Pie Chart

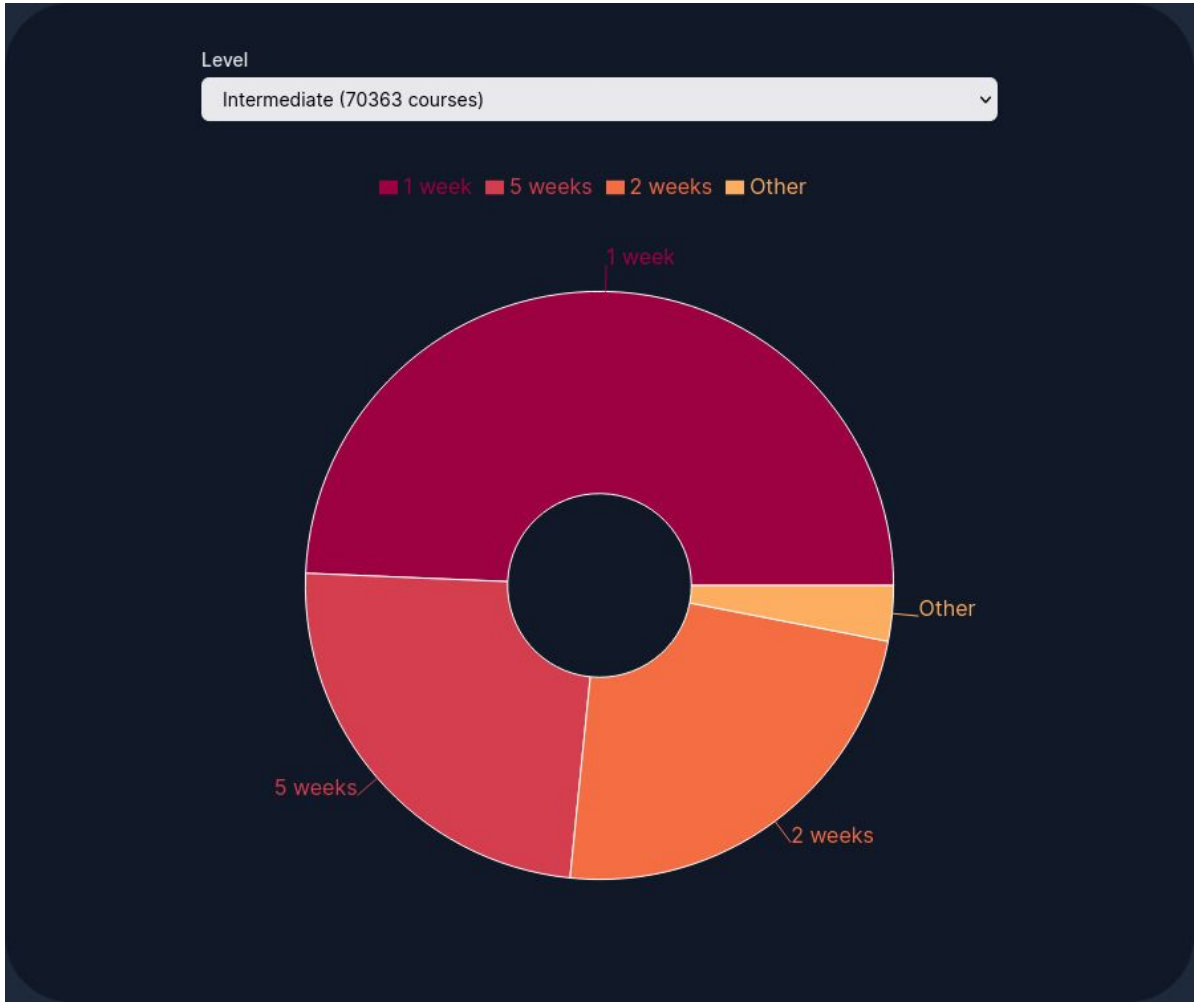
What?

- 1 Categorical key attribute (duration)
- 1 Quantitative value attribute

  (number of courses)

How?

- Marks: Area
- Channels: Tilt



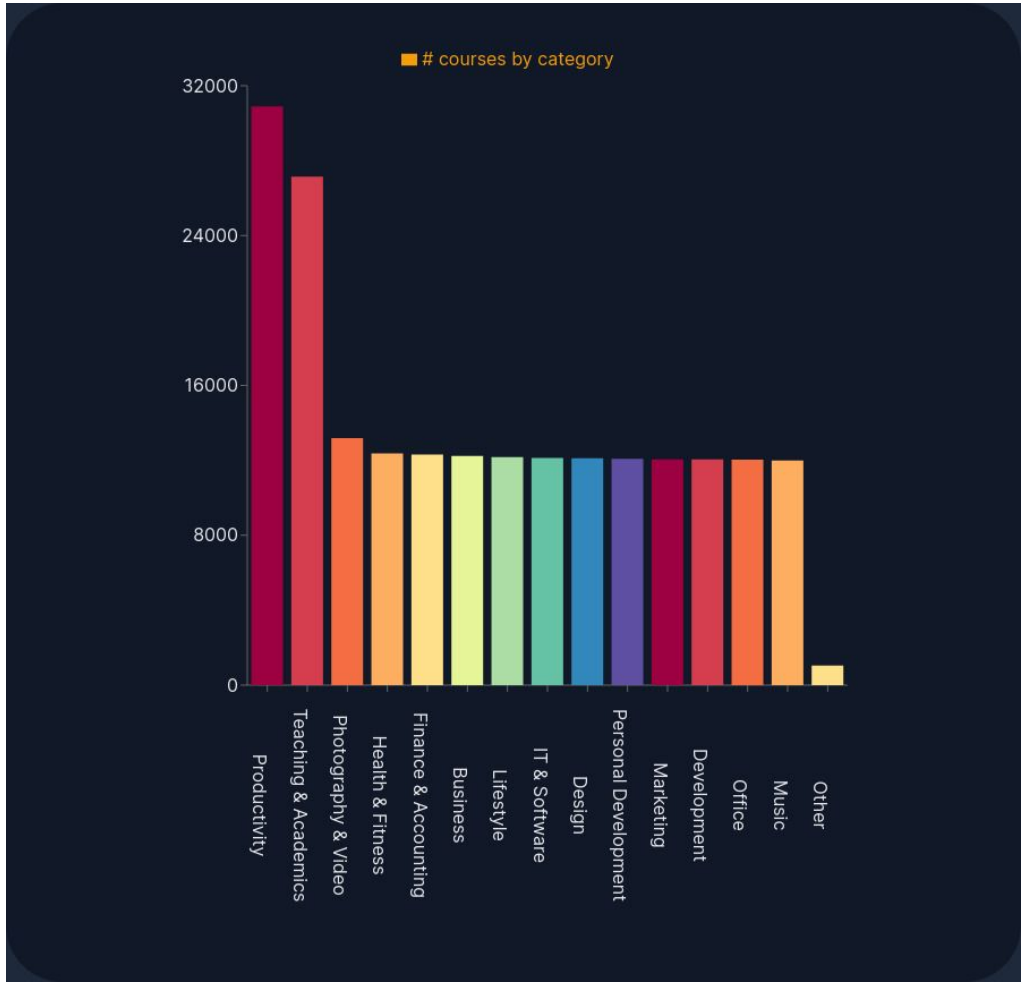
Bar Chart

What?

- 1 Categorical key attribute
- 1 Quantitative value attribute

How?

- Mark: **Lines**
- Channels:
  - Spatial regions:
    - Separated horizontally
    - Aligned vertically
    - Ordered by number of courses in the category
  - Length -> to express quant. value



Alternative design choices

How: Facet	Juxtapose
How: Reduce	Filter, Aggregate
How: Manipulate	Select, Navigate

Machine Learning Algorithm Used

Naive bayes Algorithm

- Multinomial and sparse (text classification)
- To categorise courses based on the descriptions as only half of the course are labeled
- Train data - Udemy (labeled)
- Labels the rest of the data (Coursera and EDX)

Recommender System Algorithm Used

Content based recommendation

- After we minimise the number of courses based on factors such as price & durations

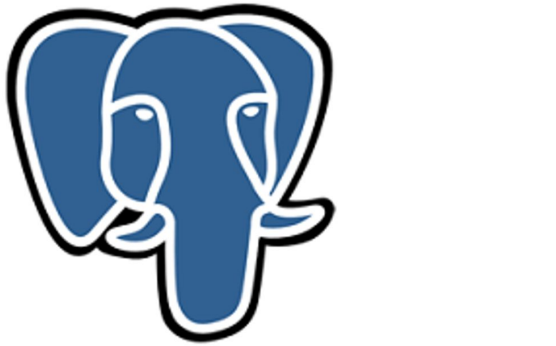
- **TF- IDF matrix** with stop words removed to quantify the descriptions
- **Elbow method** to decide the amount of clusters
- **K-means clustering** to put them into different clusters and return the top words of each clusters to the user to swipe on

Implementation - Preparation

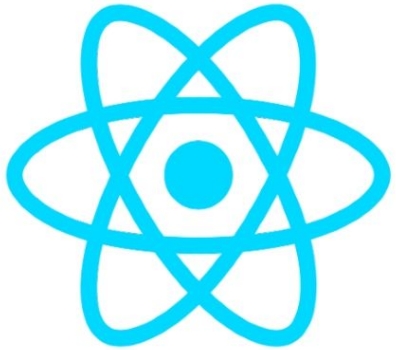
Implementation - Backend

Algorithms

Connections



Implementation - Frontend



React

Recharts

tailwindcss

Results (10 min)

Summary

What

- Data : List of Online courses offered in MOOCs (study plan)
- Environment: Online Learning

Why

- Personalisation and Recommendation

How

Methods:
K mean, TF-IDF matrix, Data mining, Information visualisation, Content based recommendation

Who

- Students: To find a suitable course for them to learn from