Udemy – An analysis of Learner's satisfaction

Presented by : Timothy Chan ZX
for LHUB Data Analytics program - Capstone Project 2

Agenda

1. **Background**
   - The Task at hand
   - Problem to solve
   - Data source
2. **Data Structure**
   - ER
   - Schema Diagram
3. **Data Analysis**
4. **Data Cleaning**
5. **Dashboard**
6. **Conclusion and Recommendation**

Task - Gain Customer Experience insights into Udemy

Background- Udemy Experiences High Growth Across Asia-Pacific Region and Melbourne Hub.

As business grows, is Customer Satisfaction keeping up?
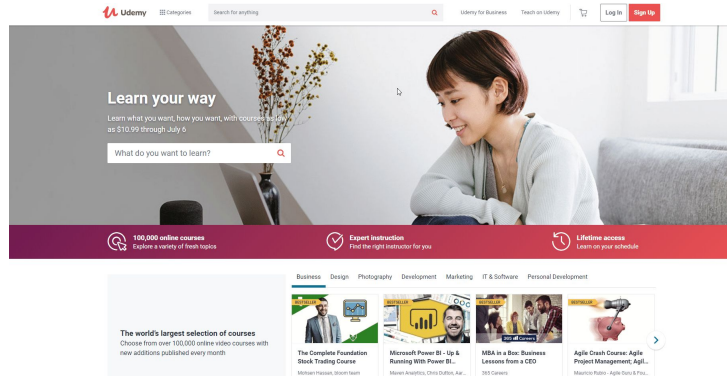
Problem statement

- As business grows, is Customer Satisfaction keeping up?
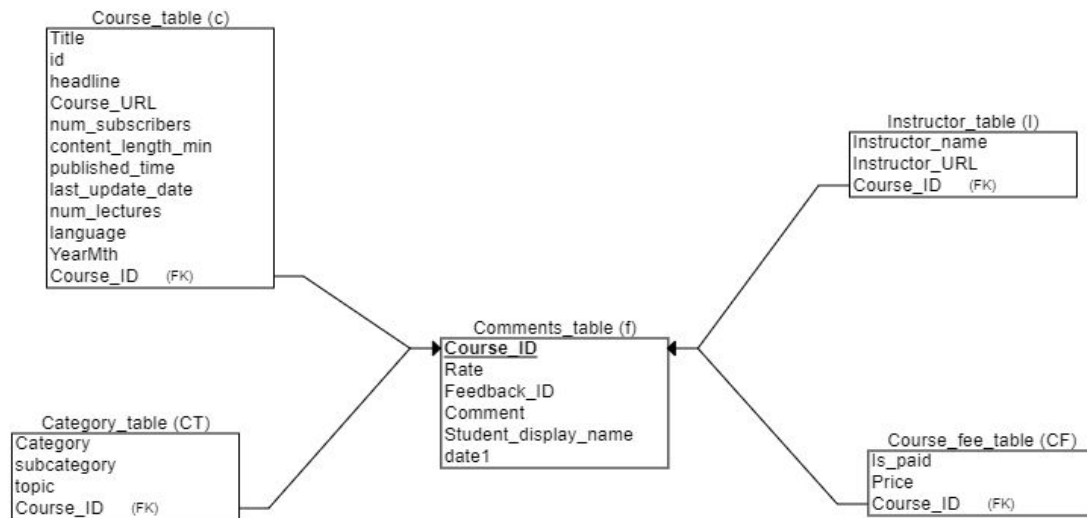
Questions to ponder -

- How satisfied are our Learners?
- What are the categories?
- Who are the good instructors?
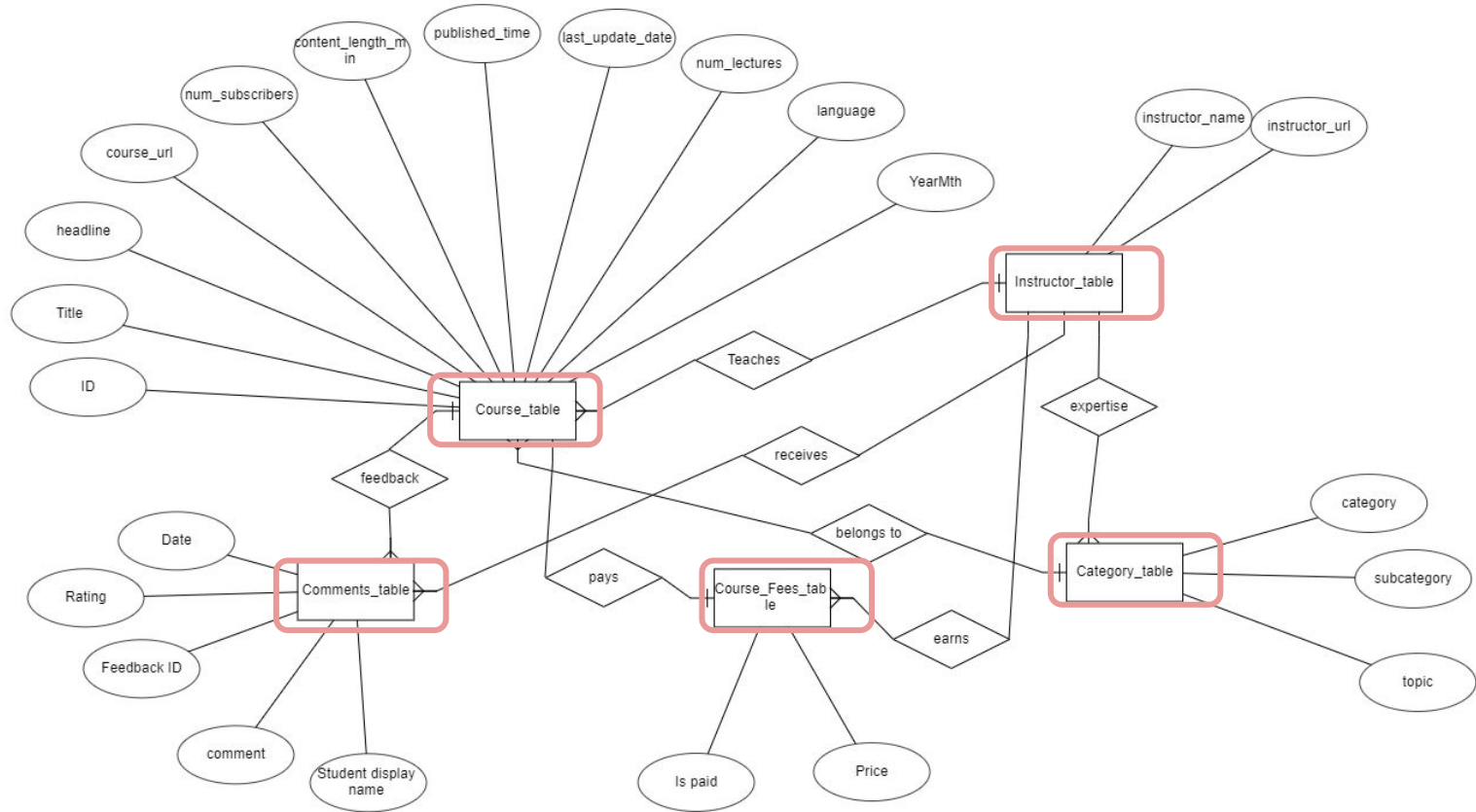- Are there any difference between paid and unpaid courses?

Data source

- https://www.kaggle.com/datasets/hossaingh/udemy-courses?select=Comments.cs
- This dataset contains detailed information on all available Udemy courses on Oct 10, 2022.
- Udemy holds 209,734 courses and 73,514 instructors teaching courses in 79 languages in 13 different categories
- Data in 2 tables
    - Comments.csv - over 9 million comments
    - Course_info.csv - 209k courses

Relational Schema

Course_table (c)
- Title
- id
- headline
- Course_URL
- num_subscribers
- content_length_min
- published_time
- last_update_date
- num_lectures
- language
- YearMth
- Course_ID   (FK)

Instructor_table (I)
- Instructor_name
- Instructor_URL
- Course_ID   (FK)

Comments_table (f)
- Course_ID
- Rate
- Feedback_ID
- Comment
- Student_display_name
- date1

Category_table (CT)
- Category
- subcategory
- topic
- Course_ID   (FK)

Course_fee_table (CF)
- Is_paid
- Price
- Course_ID   (FK)

ER Diagram

Data Analysis

Extracting rating by year of feedback

Extracting rating by year of Publication

Data Analysis

Extracting rating by
no. of feedback and
average rating for
2022

```sql
304     select left(f.date1,4) Year ,  CF.is_paid Paid, count(CF.is_paid) num_feedback, avg(f.rate) rating
305   ⊟ from Comments_table f
306     left join Course_fee_table CF
307     on (f.course_id=CF.course_id)
308     left join Category_table CT
309     on (f.course_id=ct.course_id)
310     where left(f.date1,4) = 2022
311     group by  left(f.date1,4), cf.is_paid
312     order by left(f.date1,4) asc, rating desc
313
314
```

150 %

▦ Results ▦ Messages

	Year	Paid	num_feedback	rating
1	2022	TRUE	1428305	4.63269889834454
2	2022	FALSE	149761	4.58610052016213

Data Analysis

Extracting top 5 and
bottom 5 of Course
category ratings

```sql
422
423  select top 5 ct.subcategory, count(f.rate) num_feedback, avg(f.rate) rating from Comments_table f
424  left join category_table ct
425  on (f.course_id=ct.course_id)
426  where left(f.date1,4) = 2022
427  group by ct.subcategory
428  order by avg(f.rate) desc
429
```

50 %

Results | Messages

	subcategory	num_feedback	rating
1	Esoteric Practices	19654	4.81021674977104
2	Yoga	4351	4.79728797977476
3	Meditation	1974	4.78723404255319
4	Dance	630	4.76349206349206
5	Religion & Spirituality	3893	4.75777035705112

```sql
429
430  select top 5 ct.subcategory, count(f.rate) num_feedback, avg(f.rate) rating
431  from Comments_table f
432  left join category_table ct
433  on (f.course_id=ct.course_id)
434  where left(f.date1,4) = 2022
435  group by ct.subcategory
436  order by avg(f.rate) asc
437
```

150 %

Results | Messages

	subcategory	num_feedback	rating
1	Taxes	778	4.42030848329049
2	SAP	6302	4.46921612186607
3	Science	4745	4.47081138040042
4	Fashion Design	1092	4.47252747252747
5	Video & Mobile Marketing	1338	4.47309417040359

Data Analysis

Calling a function to pass 3 variables – year and rating

Function will return the value of the CSAT score

```sql
392  select dbo.calcsat(4,5,2022) as CSATscore


370  create function calCSAT (@numlow real, @numhigh real, @Year real)
371  Returns real
372  as
373  begin
374  declare @val1 real
375  declare @val2 real
376  declare @result real
377
378  set @val1 = (select count(rate) from Comments_table
379  where rate between  @numlow and @numhigh and left(date1,4) = @year)
380
381  set @val2 = (select count(rate) from Comments_table
382  where left(date1,4) = @year)
383
384  set @result =  @val1 / @val2
385
386  return @result
387  end
```

Data Cleaning

Changing the variable types of some fields so that it can be used for calculation with other variables.

```
52
53   Alter table course_table alter column yearmth nvarchar(30)
54
55
56   Alter table Course_fee_table alter column is_paid nvarchar(30)
```

```
413  Alter table Comments_table
414  Alter Column rate real
```
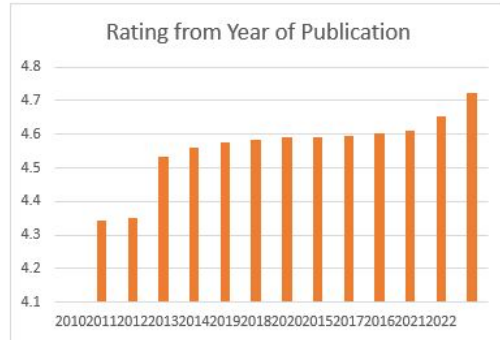
Customer Satisfaction Analysis

Ratings is from a scale of 1 to 5

Udemy Customer Experience Analysis

Total feedback in 2022 - **1,578,066**

Average Score **4.72**
CSAT Score **91.7%**

Learner's satisfaction has gone up for newer courses

Rating from Year of Publication

Course satifaction rating in 2022

Learner's highest satisfaction - ~ 0.4 pts difference between top and bottom 5

Category	No. of feedback	Rating
Esoteric Practices	19654	4.81
Yoga	4351	4.80
Meditation	1974	4.79
Dance	630	4.76
Religion & Spirituality	3893	4.76

Category	No. of feedback	Rating
Taxes	778	4.42
SAP	6302	4.47
Science	4745	4.47
Fashion Design	1092	4.47
Video & Mobile Marketing	1338	4.47

Satisfaction rating of most popular instructors - all above average of 4.72

Instructor Name	No. Feedb	Rating
Dr. Angela Yu	20889	4.77
Stephane Maarek	AWS	18097	4.80
365 Careers	14023	4.69
Jose Portilla	12853	4.68
Kyle Pew	12014	4.81

Dip in Learner's satisfaction during Covid Period

Rating for year of feedback

Breakup of Paid and Unpaid Courses in 2022

Over 90% of feedback are from Paid courses

149761 9%

1428305 91%

Paid and Upaid Courses - no significant differnece

4.59 50%   4.63 50%
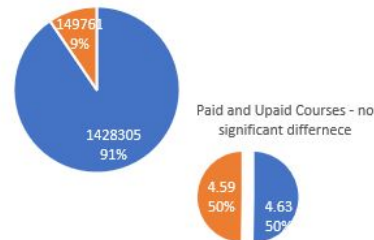
Distribution of Satisfaction score
- Majority of rating at 4 and 5.
- There are more 1s than 2s

One Two Three Four Five

Five, 1211461

Four, 234854

Thr... 761...

O...

T...
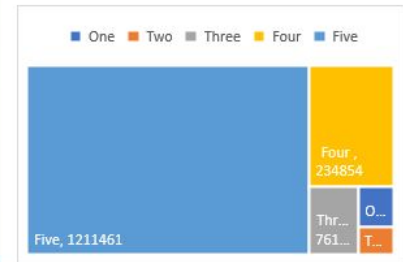
In Conclusion

- Learner's satisfaction score has been improving
- There is no significant difference between paid and unpaid courses
- Learner's satisfaction at the Category level is not significant too
- Distribution of Satisfaction score is high towards 4 and 5

Recommendation

- Deep dive into data analytics on the unstructured data from the Learners' comments for the categories with the lowest customer ratings