

**Exploratory  
Data Analysis**

**to Udemy  
Courses**



# Outline

1

introduction

2

Data Structure

3

Data Cleaning

4

Data Analysis

5

Consclusion



# Introduction

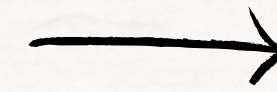
This project is for the dataset which contains all the courses that published between 2011 to 2017, It aims to analyze:

- the behavior of the courses with the highest rating and the most requested courses
- price of the courses
- number of subscribers of each subject, number of courses per subject.



## Data Structure

DataSet



3679

X

12

Observations

Features

- **Course\_id**: assigned to each course
- **Course\_title**: title for each course
- **url**: link for courses
- **is\_paid**: it is free or not?
- **price**: price per course
- **num\_subscribers**
- **num\_reviews**
- **num\_lectures**: how many lectures in the course
- **level**: course level
- **content\_duration**: How many hours in the course
- **published\_timestamp**: Dates, day and time when each course was generated
- **subject**: topic of courses



# Data Cleaning

- No Null values were found in the dataset
- Drop column **url**
- Rename all small first letters to capital letters

```
Data.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 3678 entries, 0 to 3677  
Data columns (total 12 columns):  
#   Column                Non-Null Count  Dtype    
---  ---                     -  
0   course_id             3678 non-null   int64    
1   course_title          3678 non-null   object   
2   url                   3678 non-null   object   
3   is_paid               3678 non-null   bool      
4   price                 3678 non-null   int64    
5   num_subscribers       3678 non-null   int64    
6   num_reviews           3678 non-null   int64    
7   num_lectures          3678 non-null   int64    
8   level                 3678 non-null   object   
9   content_duration      3678 non-null   float64   
10  published_timestamp    3678 non-null   object   
11  subject               3678 non-null   object   
dtypes: bool(1), float64(1), int64(5), object(5)  
memory usage: 319.8+ KB
```

No  
Null  
values



```
Data.drop('url',  
         axis='columns', inplace=True)
```

Drop url

```
Data.rename(columns={  
    'course_id': 'Course_Id',  
    'course_title': 'Course_Title',  
    'is_paid': 'Is_Paid',  
    'price': 'Price',  
    'num_subscribers': 'Num_Subscribers',  
    'num_reviews': 'Num_Reviews',  
    'num_lectures': 'Num_Hours',  
    'level': 'Level',  
    'content_duration': 'Content_Duration',  
    'published_timestamp': 'Published_Timestamp',  
    'subject ': 'Subject ',  
}, inplace=True)
```

Rename

```
Data.columns
```

```
Index(['Course_Id', 'Course_Title', 'Is_Paid', 'Price', 'Num_Subscribers',  
      'Num_Reviews', 'Num_Hours', 'Level', 'Content_Duration',  
      'Published_Timestamp', 'subject'],  
      dtype='object')
```



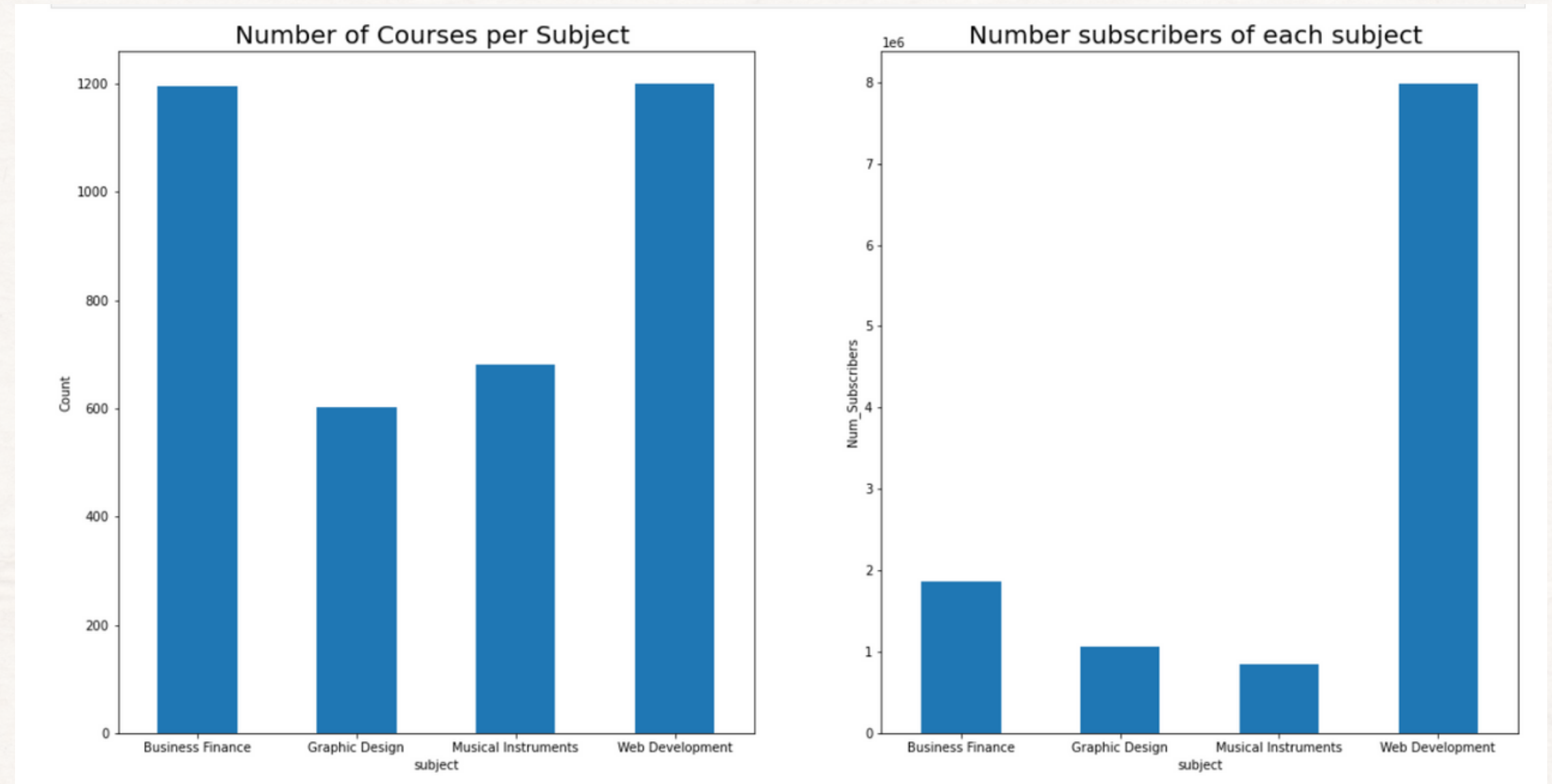
# Data Analysis

Are  
you  
ready?



# Observation

number of subscribers of each subject, number of courses per subject.



	Num Courses	Subscribers
Business Finance	1195	1868711
Graphic Design	603	1063148
Musical Instruments	680	846689
Web Development	1200	7980572



## Observation

```
A= Data['Price'].idxmin()  
B= Data['Price'].max()  
  
print("The minimum number of courses price = %d" %(A))  
print("The maximum number of courses price = %d" %(B))
```

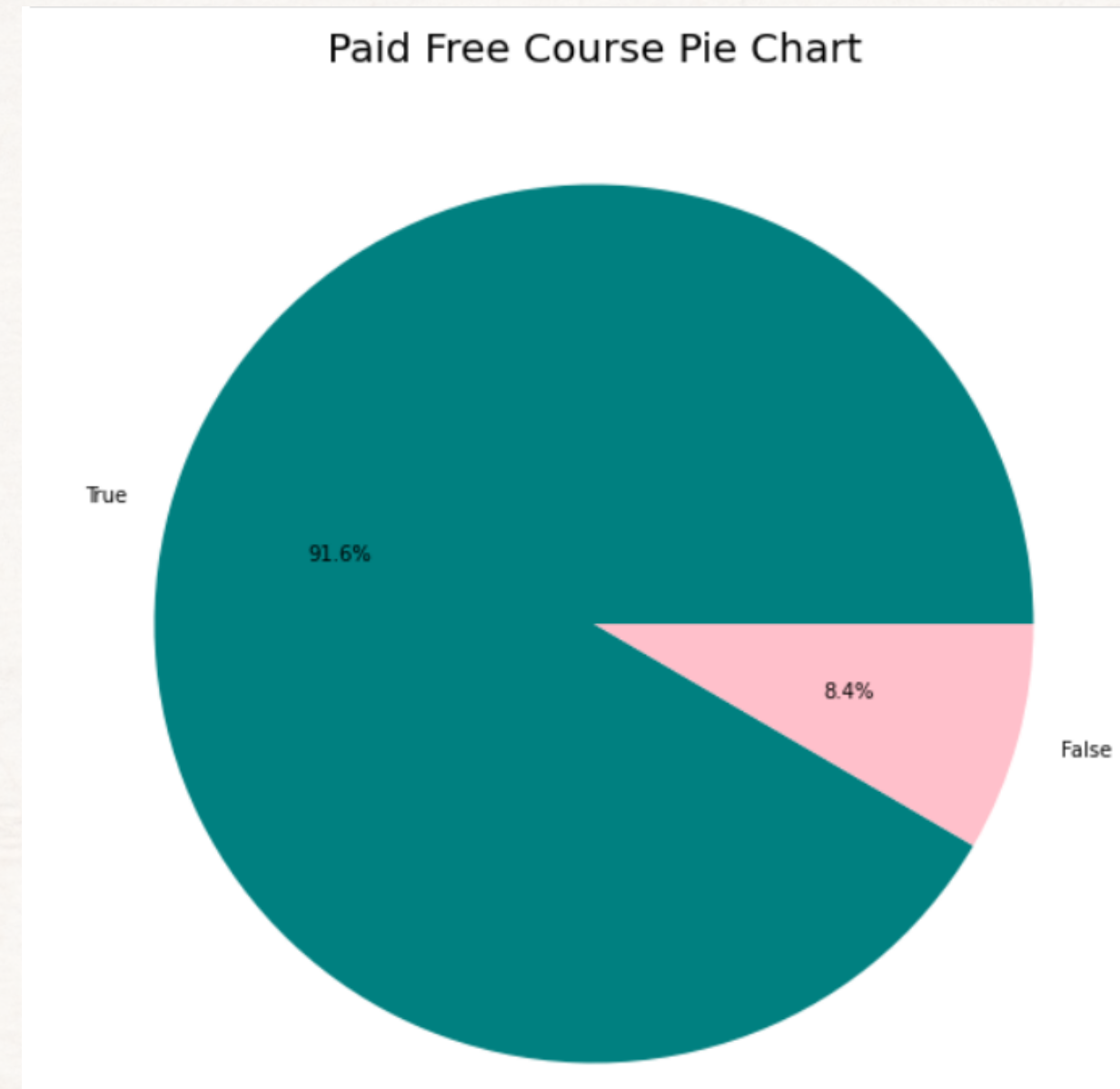
```
The minimum number of courses price = 95  
The maximum number of courses price = 200
```

The minimum and maximum of  
the price in courses



Observation

Paid courses



```
Data['Is_Paid'].value_counts()
```

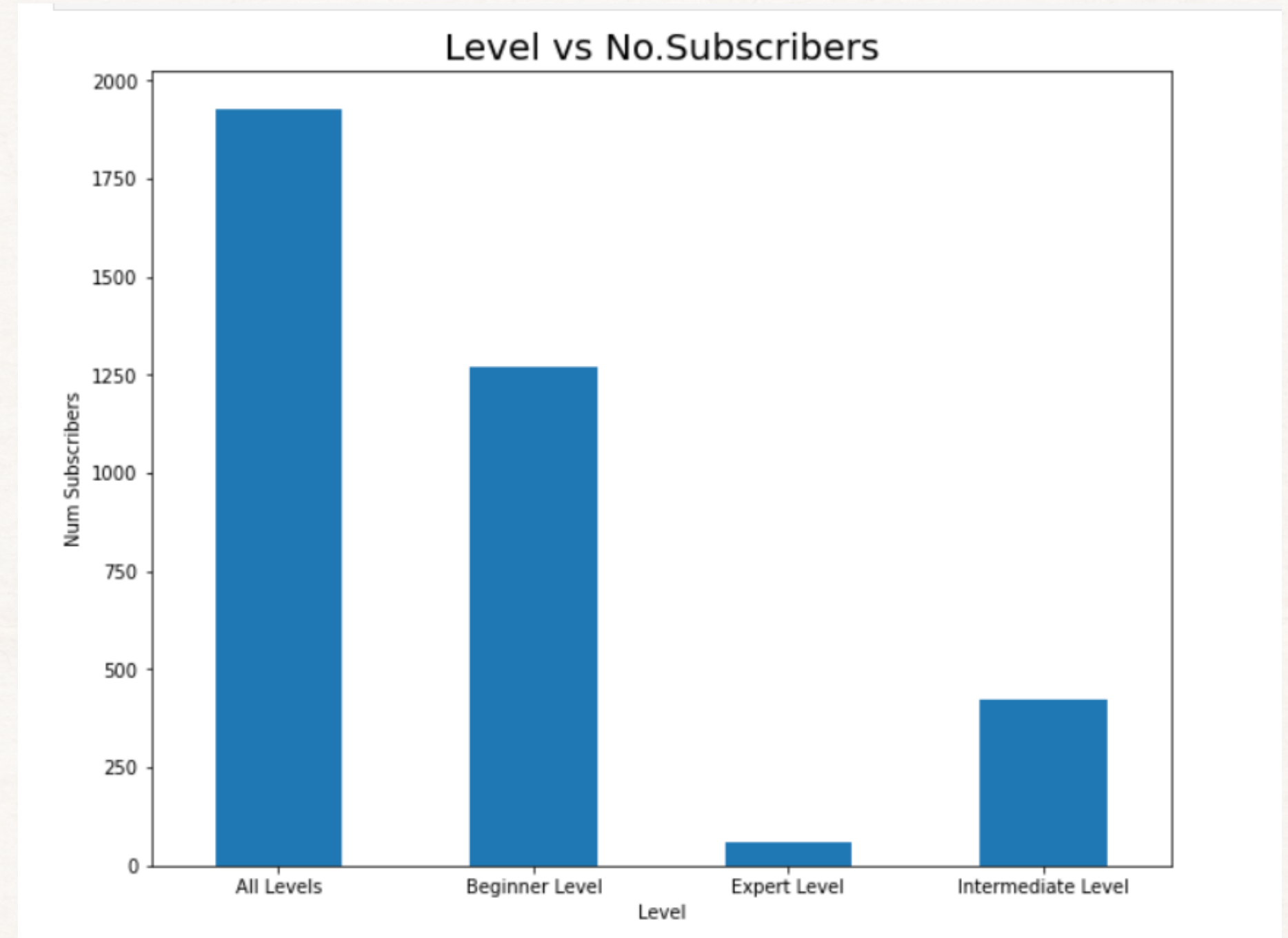
```
True      3368  
False      310  
Name: Is_Paid, dtype: int64
```



## Observation

num courses of each  
topic and num  
subscribers of each  
course

	Num Courses	Subscribers
Expert Level	58	50196
Intermediate Level	421	742005
Beginner Level	1270	4051843
All Levels	1929	6915076





# Observation

Top 5 most expensive course for each topic

Top 5 most expensive Web Development course

	Course_Title	Price	Num_Subscribers
3533	Kids Coding -Introduction to HTML, CSS and Jav...	200	2355
3620	Learning Path: The Road to Elasticsearch	200	50
3489	Ultimate guide to Social Media Web development...	200	11011
2758	Complete HTML5 and CSS3 Course +1 Start to Fin...	200	3887
2756	WordPress Tips and Tricks	200	1117

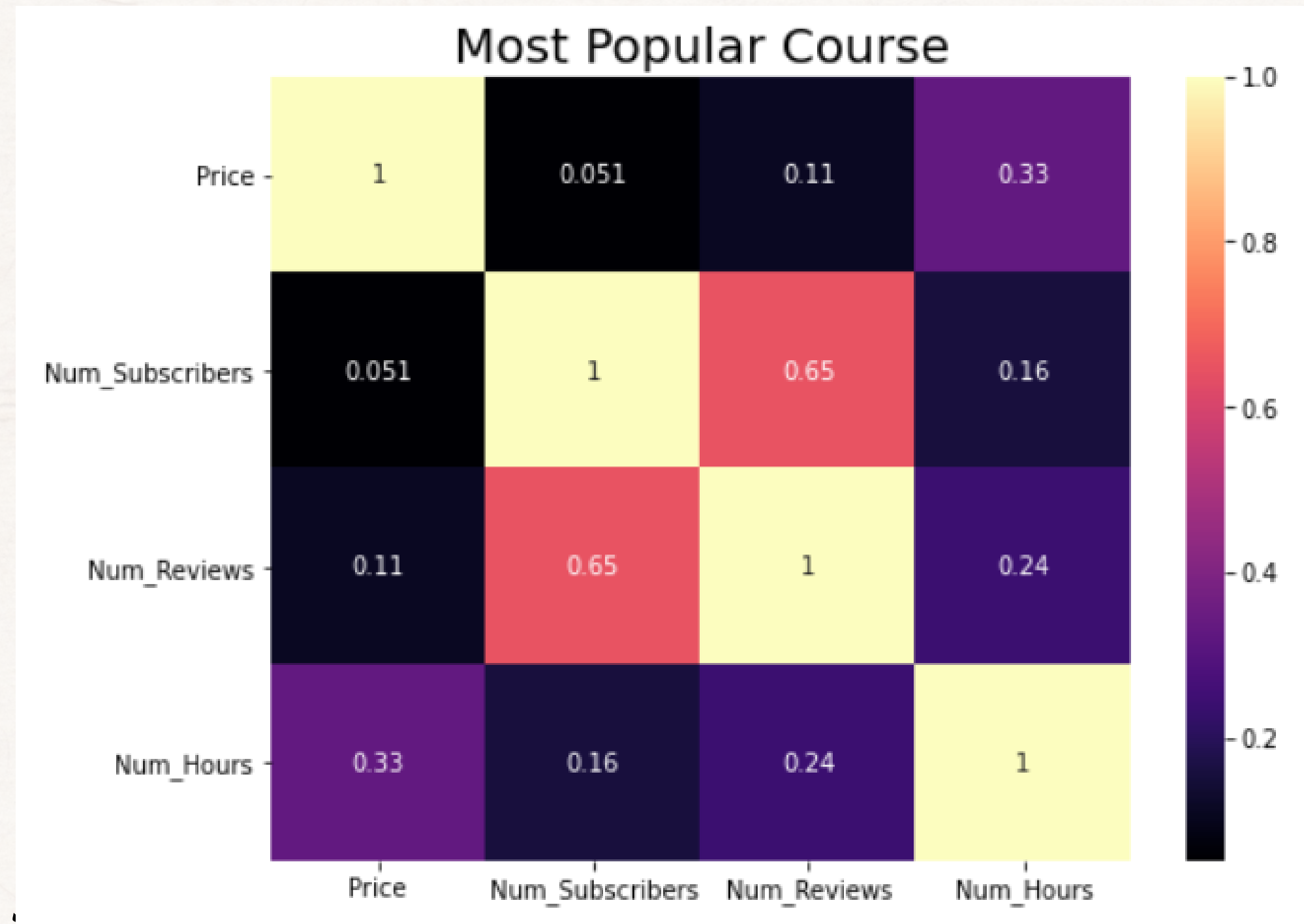
Top 5 most expensive Business Finance course

	Course_Title	Price	Num_Subscribers
0	Ultimate Investment Banking Course	200	2147
190	Investing: How I Consistently Beat The Stock M...	200	2000
147	Python Algo Trading: FX Trading with Oanda	200	453
568	Successful Day Trading with Fibonacci	200	869
538	Intro to Impact Investing	200	44



# Observation

Most popular courses





Thank  
you!

Have a  
great  
day  
ahead.