

Data Play reviews: Count vectorizer and Word cloud

NATHAN CHRISTOPHER MENON



Agenda

- ▶ Problem Statement
- ▶ Overview of the Data
- ▶ Tools Used
- ▶ Methodology: Count Vectorizer
- ▶ Result: Count vectorizer
- ▶ Methodology: Word Cloud
- ▶ Result: Word Cloud
- ▶ Inferences

Problem Statement

- ▶ Data Play is an Ed-tech company that is focused on providing quality education to its students and making them industry-ready
- ▶ Data play has received feedback on its teaching in the form of Google reviews
- ▶ The data was scrapped and imported into an Excel sheet
- ▶ The sheet consists of the Name of the person who left the review, the rating the person left, and the description of the review
- ▶ Based on that we want to know whether the crowd has positive feedback about the company or a negative one

100

[illegible]

Tools Used

- ▶ **Text-to-Columns:** Splits text based on a delimiter(space in this case)
- ▶ **Trim:** Removes extra white space before and after a text
- ▶ **Unique:** To get the unique words from the text
- ▶ **Count if:** To get the frequency of a value(unique values in this case) within a range of text
- ▶ ***We use Excel for the count vectorizer and Power BI for the wordcloud***

Methodology: Count Vectorizer

- ▶ The method used here is depicting count vectorizer (creating a dictionary consisting of the unique words and their frequency)
- ▶ This is done to clarify how the words are counted and the word cloud visualized
- ▶ First of all, we use text to columns under the data tab and split the text into multiple columns
- ▶ Then we copy the data in each column and merge it into a single column
- ▶ Using trim, we remove the extra spaces before and after the words
- ▶ Then, using unique we extract the unique words and using the count if function, specifying the range we get the frequency of occurrence

Result: Count Vectorizer

NAMES	RATING	GOOGLE REVIEW DESCRIPTION	TRIMMED	UNIQUE VALUES	FREQUENCY OF UNIQUE VALUES
Abhinav Bharat	5	I	I	I	6
Diwanshu Choudhary	5	According	According	According	1
Jatin Bhagtani	5	Nishant	Nishant	Nishant	4
Jinal Shukla	5	Altogether	Altogether	Altogether	2
Jiya	5		2	2	1
Kurakula Prashanth	4	The	The	The	43
		DataPlay			
Mradul Shrotriya	5		DataPlay	DataPlay	9
Nishita Sharma	5	I'm	I'm	I'm	2
Prit Anand	5	Highly	Highly	Highly	5
Rishabh Vaish	5	Happy	Happy	Happy	1
Tejasvi Poonia	5	Useful	Useful	Useful	1
UTPAL KUMAR RAY	5	This	This	This	5
Vaidik Asawa	5	I'm	I'm	I'm	1
Yaman Sain	5	Altogether	Altogether	Attending	1
Yash Tak	5	The	The	It	7
Arpan Banerjee	5	The	The	Great	11
Aryan Khandelwal	5	I'm	I'm	Good	16
Chandan op	5	Attending	Attending	Data	24
Kartikey Patiyar	5	I	I	Perfect	2
Naitik Parik	5	Dataplay	Dataplay	Content	2
Naman Sehra	5	Dataplay	Dataplay	Embarking	1
Paras Kaushik	5	It	It	Nice	4
Prit Anand	5	The	The	had	1
Sangam Dhaker	5	Great	Great	to	28
Satyam Kumar	5	DataPlay	DataPlay	sir	4
Vaidik Asawa	4	I	I	days	1
Vansh Mathur	5	DataPlay	DataPlay	explanations	1
Sourabh Thanvi	5	Good	Good	is	17
Rabin Sunuwar	5	Data	Data	currently	4
Rahul Joshi	5	Perfect	Perfect	recommended	2
DIVYANGAN	5	This	This	For	19
Sutesna Mondal	4	Content	Content	experience	8

Methodology: Word Cloud

- ▶ Import the Excel sheet containing the data play reviews into Power BI
- ▶ Go to the visualization tab and import Word Cloud by Microsoft
- ▶ In the data tab, select the column that contains the description of the reviews
- ▶ And..... The word cloud is generated!!
- ▶ We select the option to remove stop words as we want relevant words that reflect the reviews to appear exclusively
- ▶ ***Stop words: Connecting words like is, are, be, etc.***

100

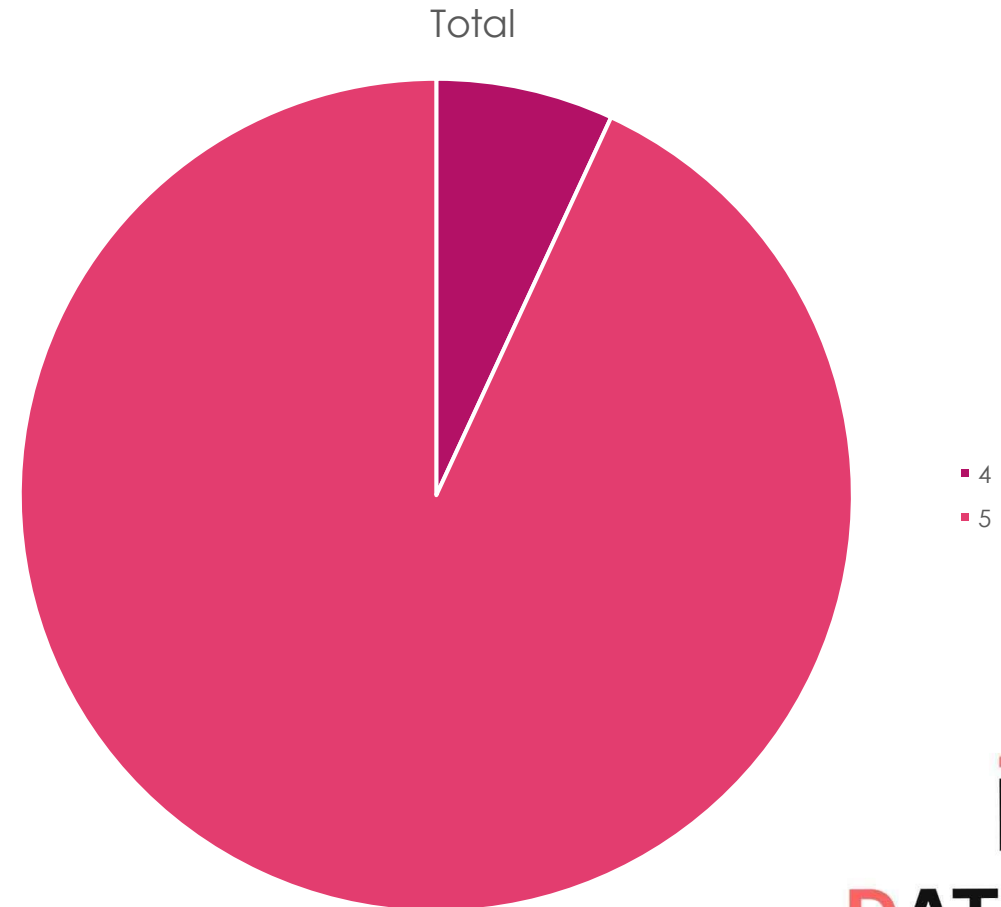


Inferences

From the word cloud, we can see that words like **Very good**, **Learning**, **experience**, **great**, **perfect**, **improve skills**, etc.

This indicates positive feedback and shows that Data Play lives up to its name

As you can see in the pie chart, it has received reviews of only 4 or 5 out of 5 with a rating of 5 dominating the pie chart, which proves my inferences about Data play



THANK YOU!