Leveraging SPoT Surveys for Instructor Evaluation

1st Almonte, Abiel
Dept. Electrical and Computer Engineering
College of Engineering and Computing (FIU)
Miami, United States
aalmo061@fiu.edu, PID: 6229783

***Abstract*—The study emphasizes the crucial role of semester course selection and instructor choice in shaping a student's academic performance. It introduces the use of big data analytics, specifically Spark, to analyze FIU Student Perceptions of Teaching survey data, transforming it into JSON format. The analysis focuses on aggregating percentages to create a comprehensive measure of instructor quality, overcoming challenges posed by diverse percentage scales and class sizes based on a proposed formula. Descriptive statistical analysis provides insights into rating distributions and the impact of response counts on ratings. The study concludes by underlining the usefulness of SPoT survey data in evaluating instructor performance and suggests potential integration into user-friendly applications to improve course selection processes.**

***Index Terms*—FIU (Florida International University), Instructor Evaluation, Big Data Analytics, Spark, Python, Descriptive Statistics, Comparative Analysis**

I. Introduction

Selecting semester courses and by extension the corresponding instructor significantly impacts a student's success in regards to GPA due to unique teaching styles and varying difficulties. A naive yet reliable approach to ensure the instructor is of good quality is by trusting the reviews and ratings published by previous students. Florida International University's (FIU) public repository for student reviews is known as FIU's Student Perceptions of Teaching (SPoT) Surveys [1]. However, SPoT surveys fail to provide a rating that is easily digestible and comparable.

Big data analytics involves analyzing extensive volumes of data to identify trends and correlations, aiding informed decision-making to solve specific problems [2]. The task is to carefully aggregate data into individual instructor ratings based on a point system that can be further analyzed to determine the best professor under specific conditions.

II. Description of Dataset

A. Overview

Located in FIU Analysis Information & Management's (AIM) web page. The SPoT survey dataset is organized into column families ([1]NoSQL), where the [2]key consist of unique class credentials and the super columns are the questions and rating classifications. The dataset is loaded into Tableau, a platform for visualizing and analyzing datasets [3]. The sample of data downloaded ranged from 2020 to present day, with a total of 448,176 rows.

B. Structure in Depth

There are several row keys that are unnecessary, and therefore will be currently ignored and subsequently removed. The dataset has the following *relevant* row keys: Course, Instructor, [3]Res/ En. The Course is composed of the class name, level, and section. The Instructor field includes the name of the faculty or staff teaching the course. The Res/En is the numerical fraction of how many students that responded to the SPoT survey over the total amount enrolled into the course.

The super columns consist of the following: Question and [No Response, Poor, Fair, Good, Very Good, Excellent]. The former are unique single answer questions –the content of which is irrelevant– given during the survey, and the latter is the list of choices given to the student to classify the instructors performance where its entries are the proportions of the total amount of responses that chose the respective classification to an associated question, represented as $\boldsymbol{p}$. The list of rating classifications will be condensed into an array $\boldsymbol{C}$ for simplicity.

TABLE I
Student Perceptions of Teaching Survey

Course	Instructor	Reps./En.	Question	$\boldsymbol{C_0}$	$\dots$	$\boldsymbol{C_n}$
Data Sci.			q_0	ρ_{00}	$\cdots$	ρ_{0n}
CAP XXXX	*Last, First*	*Res/En*	$\vdots$	$\vdots$	$\ddots$	$\vdots$
RVC			q_i	ρ_{i0}	$\cdots$	ρ_{in}

*Sample of records that represent SPoT survey results for a singular class.

Where $\boldsymbol{C_n}$ represents the rating classification at the *n^th^* index of the [4]C array, $\boldsymbol{q_i}$ is the question at the i^{th} index in an array of unique questions, and $\boldsymbol{\rho_{in}}$ is $\boldsymbol{p}$ at the i^{th} row, n^{th} column of the $i \times n$ matrix [5]$\boldsymbol{\rho}$. Where each row represents a unique question, and each column represent a different rating classification.

[1]Non-relational database
[2]Unique assortment of row keys (within this context)

[3]Abbreviation for Responses/Enrollment
[4][No Response, Poor, Fair, Good, Very Good, Excellent]
[5]$\boldsymbol{\rho}$ is an $i \times n$ matrix of all varying ***p's*** for a singular class survey

III. Dataset Processing

A. Languages

Processing requires the utilization of **Python** and Spark **SQL**. Python will be employed to originally load the dataset and create the functions necessary to execute the queries made with SQL.

B. Tools

The Python API for Apache Spark is leveraged for the analysis of the data, known as PySpark [4]. I will be utilizing the features including Spark SQL, DataFrames, and Spark RDD to preform quick, in-memory, computations on the dataset. Then finally load the JSON records into the MongoDB Atlas Data API for future utilization.

C. Techniques

The Extract Transform and Load (ETL) [5] technique is employed to drop irrelevant columns, pre-process percentages and fractions in the form of strings to numerical values, and convert the original dataset into JSONs to be stored in a Spark RDD that can be loaded as a Spark Dataframe if necessary.

The main focus of the analysis is aggregating the $\boldsymbol{\rho}$ for every class to form a holistic rating of an instructor that can be interpreted at face value and compared to other instructor ratings, because the current dataset is simply displaying a collection of percentages for one class that have no relation to the next collection for another class. The formula proposed that models the technique of aggregating to produce a rating is as follows:

$$\frac{R^2}{E}\sum_{i=0}^{Q}\left(\sum_{n=0}^{c}\omega_n \cdot \rho_{in}\right) = \gamma \tag{1}$$

where:

- R $\equiv$ Number of total responses
- E $\equiv$ Number of total possible responses (Enrollments)
- Q $\equiv$ Number of questions
- c $\equiv$ Number of rating classifications
- ω $=$ $[0, 4, 9, 16, 25, 36]$ $\equiv$ Weights associated with the rating classifications of the same index in the $\boldsymbol{C}$ array. **e.g.** $\omega_{\mathbf{3}} = \mathbf{16}$ will be the weight applied to proportion of responses for the $C_{\mathbf{3}}$ =***"Good"*** rating classification.
- ρ $\equiv$ The $i \times n$ matrix of proportions ($\boldsymbol{p}$) previously mentioned in section I.
- γ $\equiv$ Rating of an instructor for a given class.

Since (1) does not account for instructors teaching multiple classes (duplicates), their ratings are averaged. Other fields including Course, Reviews, and Enrollment are amassed on a single JSON with the sole instructor's name.

All γ is normalized utilizing the function $ln(1+\gamma)$ and scaled on a range of $(0, 5)$.

The ratings of all the instructors are analyzed and visualized by computing descriptive statistical measurements.

IV. Results

A. Overview

The presentation of results are produced on the Python libraries Matplotlib and Seaborn.

B. Rating Distribution

6099 instructor ratings were produced from the original 400k+ records. A histogram depicting the distribution of ratings is shown in Fig. 1. Retaining a mean of 3.16 and a standard deviation of 0.8.

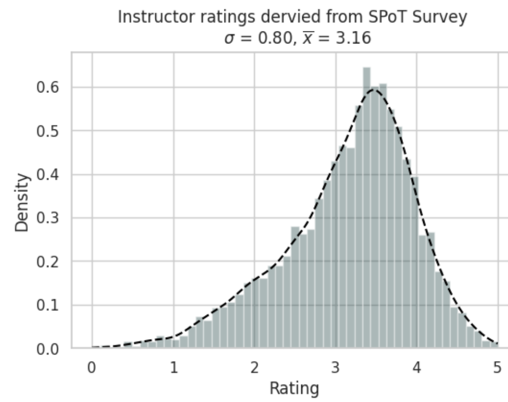


Fig. 1. Histogram of instructor ratings.

C. Review Distribution

There was a total of 800k+ responses accumulated from all the records in the dataset. Reviews is the total amount of responses an instructor received in all their surveys accounted for. Fig. 2 is a histogram displaying such distribution among all instructors.

*Note: The x-axis is truncated at 700 for visualization purposes.

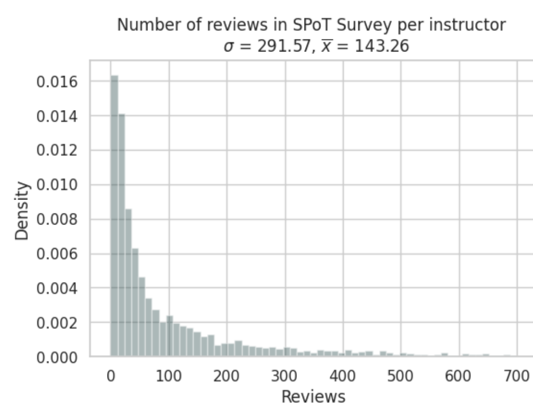


Fig. 2. Histogram of response count.

D. Impact of Responses

The response ratio has a direct correlation, clearly derived in (1). Ensuring that all instructors despite of possible class

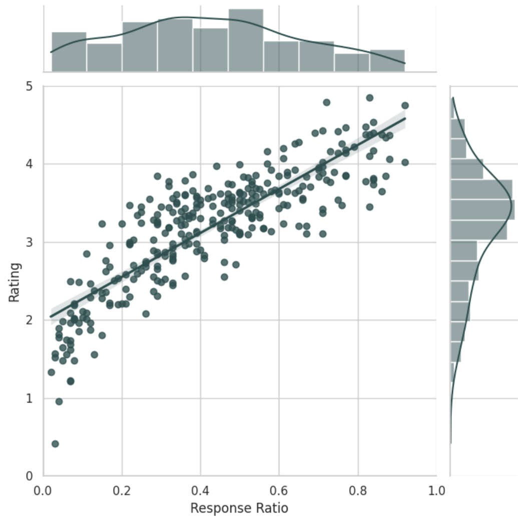


Fig. 3. Scatter plot of the response ratio's influence on ratings.

size have a fair chance to achieve similar ratings. Fig. 3 entails a scatter plot that showcases such relationship.

As for the number of responses, its influence on ratings quickly drops off, resulting in zero correlation after a trivial amount. Since a very low (< 50) amount responses will more than likely imply a poor response ratio. This trend is shown in Fig. 4.

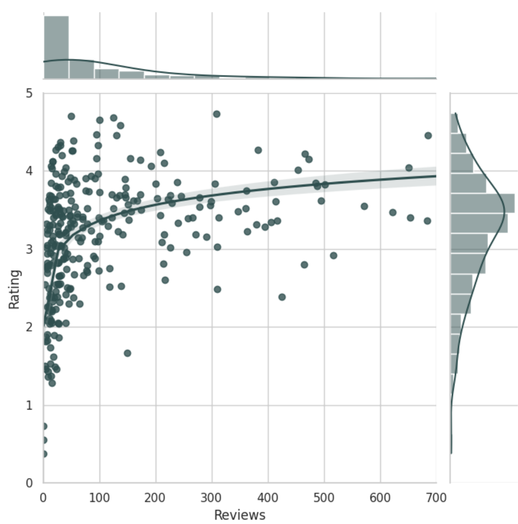


Fig. 4. Scatter plot of the response count's influence on ratings.

*Note: Fig. 3 and Fig. 4 have a random sample of data for visualization purposes.

E. Trend

It became quickly apparent and surprising, that the higher rated professors are usually the ones that are strictly an In-Person modality as shown in Fig. 5. This may be because professors in person can better encourage their students to fill out the SPoT survey, thus resulting in a higher response ratio.

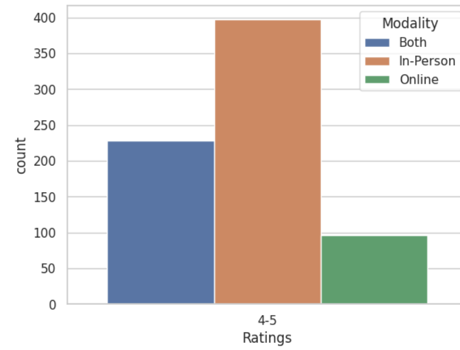
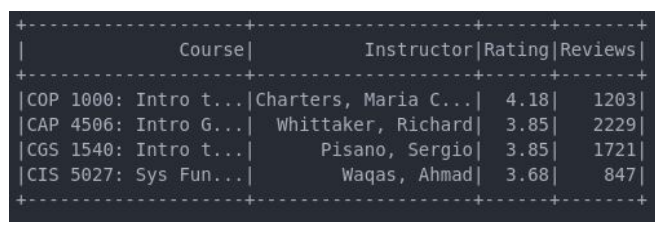


Fig. 5. Modality count in top ratings

F. Sample Practical Use

When choosing professors for courses one may want to see the best rated professor for a single course. Fig. 6 shows the result of a query that finds the top 4 rated instructors that have taught Programming 1 (COP 2210) with a reliable amount of reviews (≥ 100). The following query is employed to produce the dataframe generated in Fig. 6.

```
df= spark.read.json(jsonRDD)
df.createOrReplaceTempView('research')
sqlDF= spark.sql(
    "SELECT Course, Instructor, Rating, Reviews FROM
     research WHERE Course Like '%COP 2210%' AND
    Reviews >=100 ORDER BY Rating DESC LIMIT 4")
sqlDF.show()
```



Course	Instructor	Rating	Reviews
COP 1000: Intro t...	Charters, Maria C...	4.18	1203
CAP 4506: Intro G...	Whittaker, Richard	3.85	2229
CGS 1540: Intro t...	Pisano, Sergio	3.85	1721
CIS 5027: Sys Fun...	Waqas, Ahmad	3.68	847

Fig. 6. Result of the Spark SQL query aforementioned.

*Note: The column "Course" includes several unique courses the professor has instructed.

V. Summary

FIU AIM's SPoT survey dataset provides performance evaluations for professors in terms of percentages based on student feedback. Therefore, it is utilized as a foundation to derive ratings for instructors. By analyzing the raw data from the surveys, the aggregation formula (1) is developed to assign such ratings that enables viewers to compare professors in practical scenarios. Future steps include improving the formula for ratings and integrating the new dataset into an application that facilitates the process of course selection, whether that be a web app that provides a user interface to query, or Large Language Model that ingest the JSON documents when messaged by a user.

REFERENCES

[1] FIU Student Perceptions of Teaching Survey. (2012). opir.fiu.edu. https://opir.fiu.edu/faculty-evaluations.html

[2] Tableau. (2023). Big Data Analytics: Transform terabytes into insights. Tableau. https://www.tableau.com/learn/articles/big-data-analytics

[3] Tableau. (2021). What is Tableau? Tableau. https://www.tableau.com/why-tableau/what-is- tableau

[4] Apache Spark. (2024, February 24). Welcome to Spark Python API Docs! — PySpark 2.4.5 documentation. Spark.apache.org. https://spark.apache.org/docs/latest/api/python/index.html.

[5] IBM. (n.d.). What is ETL (Extract, Transform, Load)? — IBM. Retrieved March 30, 2024, from www.ibm.com/topics/etl website: https://www.ibm.com/topics/etl