**Data Preparation**



**1. Introduction:**

I’m trying to analyse the variables that influence a university course evaluations. As a part of

this study I’m evaluating impact of the following on course evaluation:

 **Class Size**

 **Attractiveness of staff**

 **Age of staff**

I’m carrying out the preliminary investigation of whether teaching evaluations are influenced by the attractiveness of teaching staff. This investigation is ran against the data collected by Daniel Hamermesh and Amy Parker, at University of Texas at Austin, as part of their research [1]. The data contain information about 455 course from various departments. The course evaluation rating falls between **1 ‘very unsatisfactory’** to **5**

**‘excellent’**, while the attractiveness is rated between **1 ‘lowest’** to **10 ‘highest’**. The attractiveness score was normalized to have a mean score of 0 and stored in ‘beauty’ column. All the data is stored on **‘profEvaluations.csv’** file.

**2. Relation Schema:**

After studying the relation schema I decided upon the following tables.

**Teachers** (Staffid, Age, Gender, Tenuretrack, Nonenglish, Beauty)

**Course** (Id, Staffid\*, Courseevaluation, Students, Division)

**3. Importing Data and Creating Tables: Data Import -**

 **Teachers table**

 Navigate to the established connection and right click on ‘Tables’ option.

Select ‘Import Data’ option.

 Browse and select the profEvaluations.csv file. Click next.

 In the import data window, provide the table name, in this case ‘Teachers’.

Let the import method be ‘Insert’. Click next

 In the ‘Choose Column’ window select the columns corresponding to the

above designed schema. Click next

 In the ‘Column Definition’ window iterate through each column and alter its

data-type and/or size depending on the schema and use case.

 Click finish. A new table with the above mentioned schema will be created.

 **Course table**

 Perform the same steps mentioned above. Provide the name of table as

‘Course’.

 In the ‘Column Definition’ window iterate through each column and alter its

data-type and/or size depending on the schema and use case.

 Click finish. A new table with the above mentioned schema will be created.

**Altering Imported Table -**

Now we have the tables created, we need to modify it to fit according to schema.

 **Teachers table**

 First check for any duplicate records, if they exists then remove the duplicate records.

 Next step is to set the primary key

 **Course Table**

 Similarly we check for duplicate records in Course table and set the primary and foreign key.

**ANALYSIS –**



 **Course Sizes – Number of Students**

The following table contains the minimum, mean and maximum number of students in course.

Minimum Mean Maximum

Number of students 1 3.583 20

 **Course Sizes – Course Evaluation Score**

The following table documents the minimum, mean and maximum course evaluation score provided by students, grouped by the age of the staff members. We can see that as the age group increases the mean course evaluation course decreases, hinting an inverse relation between the two variables.

Course size 18 or less 19-28 29-60 61 or more

Number of courses in group

Minimum course evaluation score Mean course evaluation score Maximum course evaluation score

112 114 116 113

2.3 2.7 2.1 2.8

4.154 4.001 3.937 3.905

5 5 5 4.8

 **Division**

The following table documents the minimum, mean and maximum course evaluation score grouped by the division of students. We can see that majority of course, 300, belongs to upper division i.e third and fourth year courses but the course in lower division tends to receive a better evaluation score, pointed by minimum score of 2.5 and mean score of 4.092

No. courses in group

Minimum Mean Maximum

Upper division 300 2.1 3.951 5

Lower division 155 2.5 4.092 5

 **Gender – Course Evaluation Score**

The following table documents the minimum, mean and maximum course evaluation score grouped by the gender of staff. We can see that on an average male staff members tends to get better score than their female counterparts, with a mean score of 4.073

No. courses in group

Minimum Mean Maximum

Female 193 2.3 3.897 4.9

Male 262 2.1 4.073 5

 **Gender – Beauty**

The following table documents the minimum, mean and maximum beauty (attractiveness) score of staff grouped by the gender of staff. We can see that female staff on an average female tends to receive better beauty score with mean beauty score of 0.125.

No. academics in group

Minimum Mean Maximum

Female 39 -1.539 0.125 1.882

Male 51 -1.511 -0.115 1.686

 **Tenure track**

The following table documents the minimum, mean and maximum course evaluation score grouped by tenure of the staff. We can see that the staff members who don’t have tenure scored a better in course evaluation, this is evident by their greater minimum (2.8) and mean (4.134) course evaluation score.

No. academics in group

Minimum Mean Maximum

Tenure track 75 2.1 3.960 5

Not Tenure track 15 2.8 4.134 5

 **Education Background**

The following table documents the minimum, mean and maximum course evaluation score grouped by whether the staff completed their undergraduate education in an English or a non- English speaking country. We can see that staff who graduated in a English speaking country tend to fair better in their course evaluation score, with a mean score of 4.102 and max score of

5, compare to mean of 3.689 and max score of 4.6 for staff with non-English speaking background.

No. academics in group

Minimum Mean Maximum

English education 83 2.1 4.012 5

Non - English education

7 2.7 3.689 4.6

 **Interactions between Tenure Track, Gender and Education Background**

In the following table we are analysing the correlation between the, gender, tenure and education background of staff, and course evaluation score they revived. We can see that male staff without a tenure and from English speaking background have the highest mean course evaluation score of 4.3961

Tenure track Gender Education No. academics in group

Tenure track Female English 28 3.928

Tenure track Female Non - English 3 3.717

Tenure track Male English 40 4.020

Tenure track Male Non - English 4 3.669

Mean

Not tenure track Female English 8 3.86

Not tenure track Female Non - English 0 N/A Not tenure track Male English 7 4.396

Not tenure track Male Non - English 0 N/A

 **Correlation Analysis**

In the following table we are tabulating the result of correlation analysis we conducted between various variable. We can infer that there is a negative relation between ‘Course evaluation score and course size’ and ‘Staff age and beauty’, -0.17286 and -0.33099 respectively, but the two sided significance is quite low < 0.05. Hence, we can say that the correlation isn’t statically significant.

The two-sided significance for ‘Staff age and mean course evaluation score’ and ‘Staff beauty and mean course evaluation score’ is > 0.05, hence the correlation coefficient for both is statically significant.

**Variables Correlation Coefficient Two-sided Significance**

Course evaluation score &

course size

-0.17286 0.00021

Staff age & beauty -0.33099 0.00144

Staff age and mean course evaluation

score

Staff beauty and mean course evaluation score

-0.02214 0.83594

0.17851 0.0923

**Discussion and Conclusions:**



From the analysis conducted above following were the **statically significant findings**:

 A positive correlation between attractiveness of staff and mean evaluation score. So the more attractive the staff member is, the better is chance of getting good course evaluation score.

 A negative correlation between ages of staff and mean evaluation score. So the younger the staff is, higher is the likelihood of the course receiving a good evaluation score.

 Male staff without a tenure and from English speaking background received the highest mean course evaluation score of 4.3961, for the respective course they teach.

**Limitation:**

 The above experiment establishes correlation between the various variable but it does establishes causation. We can never state for certain that change in one variable will

definitely impacts the other variable used in establishing the correlation.

**Reference –**



 D. Hamermesh, and A. Parker, “Beauty in the classroom: instructors’ pulchritude and

putative pedagogical productivity,” Economics of Education Review, vol. 24, pp. 369-376,

2005.

 Heckman, E., 2020. *Correlation: What It Shows You (And What It Doesn't)*. [online] Blog.minitab.com. Available at: <[https://blog.minitab.com/blog/starting-out-with-](https://blog.minitab.com/blog/starting-out-with-statistical-software/correlation-what-it-shows-you-and-what-it-doesnt) [statistical-software/correlation-what-it-shows-you-and-what-it-doesnt> [Accessed 9](https://blog.minitab.com/blog/starting-out-with-statistical-software/correlation-what-it-shows-you-and-what-it-doesnt)

June 2020].

**Appendix –**



**S.No Question SQL Queries**

1 Check ‘Teachers’ table

for duplicate records

SELECT STAFFID, COUNT(\*) FROM TEACHERS HAVING COUNT(\*) > 1

GROUP BY STAFFID;

2 Removing duplicate records from ‘Teachers’ table

DELETE TEACHERS T1

WHERE ROWID > (

SELECT MIN(ROWID) FROM TEACHERS T2

WHERE T2.STAFFID = T1.STAFFID);

3 Setting ‘staffid’ as primary key for

‘Teachers’ table

4 Check ‘Course’ table for

duplicate records

ALTER TABLE TEACHERS

ADD CONSTRAINT STAFF\_PK PRIMARY KEY (STAFFID);

SELECT ID, COUNT(\*) FROM COURSE

HAVING COUNT(\*) > 1

GROUP BY ID;

5 Setting up primary and foreign key for Course table

ALTER TABLE COURSE

ADD CONSTRAINT ID\_PK PRIMARY KEY (ID);

ALTER TABLE COURSE

ADD CONSTRAINT FK\_STAFF FOREIGN KEY (STAFFID) REFERENCES TEACHERS(STAFFID);

6 Course Sizes – Number of Students

SELECT MAX(ENROLLED), AVG(ENROLLED), MIN(ENROLLED)

FROM (SELECT COUNT(STUDENTS) AS ENROLLED FROM COURSE

GROUP BY STUDENTS);

7 Course Sizes – Course

Evaluation Score

SELECT COUNT(STUDENTS), MAX(COURSEEVALUATION), AVG(COURSEEVALUATION), MIN(COURSEEVALUATION)

FROM COURSE

WHERE STUDENTS < 19;

SELECT COUNT(STUDENTS), MAX(COURSEEVALUATION), AVG(COURSEEVALUATION), MIN(COURSEEVALUATION)

FROM COURSE

WHERE STUDENTS BETWEEN 19 AND 28;

SELECT COUNT(STUDENTS), MAX(COURSEEVALUATION), AVG(COURSEEVALUATION), MIN(COURSEEVALUATION)

FROM COURSE

WHERE STUDENTS BETWEEN 29 AND 60;

SELECT COUNT(STUDENTS), MAX(COURSEEVALUATION), AVG(COURSEEVALUATION), MIN(COURSEEVALUATION)

FROM COURSE

WHERE STUDENTS > 60;

8 Division SELECT COUNT(COURSEEVALUATION), MAX(COURSEEVALUATION), AVG(COURSEEVALUATION), MIN(COURSEEVALUATION)

FROM COURSE

WHERE UPPER(DIVISION) = 'U';

SELECT COUNT(COURSEEVALUATION), MAX(COURSEEVALUATION), AVG(COURSEEVALUATION), MIN(COURSEEVALUATION)

FROM COURSE

WHERE UPPER(DIVISION) = 'L';

9 Gender – Course

Evaluation Score

SELECT COUNT(C.COURSEEVALUATION), MAX(C.COURSEEVALUATION), AVG(C.COURSEEVALUATION), MIN(C.COURSEEVALUATION)

FROM COURSE C

JOIN TEACHERS T ON C.STAFFID = T.STAFFID

WHERE UPPER(T.GENDER) = 'F';

SELECT COUNT(C.COURSEEVALUATION), MAX(C.COURSEEVALUATION), AVG(C.COURSEEVALUATION), MIN(C.COURSEEVALUATION)

FROM COURSE C

JOIN TEACHERS T ON C.STAFFID = T.STAFFID WHERE UPPER(T.GENDER) = 'M';

10 Gender – Beauty SELECT COUNT(T.BEAUTY), MAX(T.BEAUTY), AVG(T.BEAUTY), MIN(T.BEAUTY)

FROM TEACHERS T

WHERE UPPER(T.GENDER) = 'F';

SELECT COUNT(T.BEAUTY), MAX(T.BEAUTY), AVG(T.BEAUTY), MIN(T.BEAUTY)

FROM TEACHERS T

WHERE UPPER(T.GENDER) = 'M';

11 Tenure track SELECT COUNT(DISTINCT(T.STAFFID)), MAX(C.COURSEEVALUATION), AVG(C.COURSEEVALUATION), MIN(C.COURSEEVALUATION)

FROM COURSE C

JOIN TEACHERS T ON C.STAFFID = T.STAFFID WHERE T.TENURETRACK = 1;

SELECT COUNT(DISTINCT(T.STAFFID)), MAX(C.COURSEEVALUATION), AVG(C.COURSEEVALUATION), MIN(C.COURSEEVALUATION)

FROM COURSE C

JOIN TEACHERS T ON C.STAFFID = T.STAFFID WHERE T.TENURETRACK = 0;

12 Education Background SELECT COUNT(DISTINCT(T.STAFFID)), MAX(C.COURSEEVALUATION), AVG(C.COURSEEVALUATION), MIN(C.COURSEEVALUATION)

FROM COURSE C

JOIN TEACHERS T ON C.STAFFID = T.STAFFID

WHERE T.NONENGLISH = 0;

SELECT COUNT(DISTINCT(T.STAFFID)), MAX(C.COURSEEVALUATION), AVG(C.COURSEEVALUATION), MIN(C.COURSEEVALUATION)

FROM COURSE C

JOIN TEACHERS T ON C.STAFFID = T.STAFFID WHERE T.NONENGLISH = 0;

13 Interactions between Tenure Track, Gender and Education Background

SELECT COUNT(DISTINCT(T.STAFFID)), AVG(C.COURSEEVALUATION)

FROM COURSE C

JOIN TEACHERS T ON C.STAFFID = T.STAFFID WHERE T.TENURETRACK = 1

AND UPPER(T.GENDER) = 'F'

AND T.NONENGLISH = 0 ;

SELECT COUNT(DISTINCT(T.STAFFID)), AVG(C.COURSEEVALUATION)

FROM COURSE C

JOIN TEACHERS T ON C.STAFFID = T.STAFFID

WHERE T.TENURETRACK = 1

AND UPPER(T.GENDER) = 'F' AND T.NONENGLISH = 1 ;

SELECT COUNT(DISTINCT(T.STAFFID)), AVG(C.COURSEEVALUATION)

FROM COURSE C

JOIN TEACHERS T ON C.STAFFID = T.STAFFID WHERE T.TENURETRACK = 1

AND UPPER(T.GENDER) = 'M' AND T.NONENGLISH = 0 ;

SELECT COUNT(DISTINCT(T.STAFFID)), AVG(C.COURSEEVALUATION)

FROM COURSE C

JOIN TEACHERS T ON C.STAFFID = T.STAFFID WHERE T.TENURETRACK = 1

AND UPPER(T.GENDER) = 'M'

AND T.NONENGLISH = 1 ;

SELECT COUNT(DISTINCT(T.STAFFID)), AVG(C.COURSEEVALUATION)

FROM COURSE C

JOIN TEACHERS T ON C.STAFFID = T.STAFFID WHERE T.TENURETRACK = 0

AND UPPER(T.GENDER) = 'F' AND T.NONENGLISH = 0 ;

SELECT COUNT(DISTINCT(T.STAFFID)), AVG(C.COURSEEVALUATION)

FROM COURSE C

JOIN TEACHERS T ON C.STAFFID = T.STAFFID WHERE T.TENURETRACK = 0

AND UPPER(T.GENDER) = 'F'

AND T.NONENGLISH = 1 ;

SELECT COUNT(DISTINCT(T.STAFFID)), AVG(C.COURSEEVALUATION)

FROM COURSE C

JOIN TEACHERS T ON C.STAFFID = T.STAFFID

WHERE T.TENURETRACK = 0

AND UPPER(T.GENDER) = 'M' AND T.NONENGLISH = 0 ;

SELECT COUNT(DISTINCT(T.STAFFID)), AVG(C.COURSEEVALUATION)

FROM COURSE C

JOIN TEACHERS T ON C.STAFFID = T.STAFFID WHERE T.TENURETRACK = 0

AND UPPER(T.GENDER) = 'M' AND T.NONENGLISH = 1 ;

14 Correlation Analysis SELECT

ROUND(CORR\_S( COURSEEVALUATION ,

STUDENTS, 'COEFFICIENT'),5) COEFFICIENT, ROUND(CORR\_S( COURSEEVALUATION ,

STUDENTS, 'TWO\_SIDED\_SIG'),5)

TWO\_SIDED\_P\_VALUE FROM COURSE;

SELECT

ROUND(CORR\_S( AGE , BEAUTY,

'COEFFICIENT'),5) COEFFICIENT, ROUND(CORR\_S( AGE , BEAUTY,

'TWO\_SIDED\_SIG'),5) TWO\_SIDED\_P\_VALUE FROM TEACHERS;

SELECT

ROUND(CORR\_S(AGE ,

AVG(COURSEEVALUATION), 'COEFFICIENT'),5) COEFFICIENT,

ROUND(CORR\_S(AGE ,

AVG(COURSEEVALUATION), 'TWO\_SIDED\_SIG'),5)

TWO\_SIDED\_P\_VALUE

FROM COURSE NATURAL JOIN TEACHERS GROUP BY AGE , STAFFID;

SELECT

ROUND(CORR\_S(BEAUTY ,

AVG(COURSEEVALUATION), 'COEFFICIENT'),5) COEFFICIENT,

ROUND(CORR\_S(BEAUTY , AVG(COURSEEVALUATION), 'TWO\_SIDED\_SIG'),5) TWO\_SIDED\_P\_VALUE

FROM COURSE NATURAL JOIN TEACHERS

GROUP BY BEAUTY, STAFFID;