

S3827657- Database Assignment 3

1. Data preparation –

- a. Purpose of Analysis – Primary goal behind analysing the data is to observe if the evaluation scores for professors is affected by factors such as beauty, age and gender.
- b. Relational Schemas –
Staff (StaffID, Age, Gender, TenureTrack, NonEnglish, Beauty)
Course (CourseID, StaffID*, CourseEval, StudentCount, Division)
- c. Importing csv to Oracle – Importing of CSV to oracle was carried out in 2 steps
 - i. Step 1 – Data in csv file was split into two sheets – course and staff. All the staff details were copied from the csv and put into a new sheet. Duplicates were removed to get distinct staff data
 - ii. Step 2 – Tables were created in the oracle database (refer appendix) based on the schema designed in part b
 - iii. Step 3 – Right click on the table name displayed in the navigation pane on the left and select “Import Data”
 - iv. Step 4 – Select insert as import method, browse for the csv file and properly map the excel columns to the database table columns. Make sure the data type in both cases match to avoid errors while importing.

2. Data Analysis –

- a. Course size – Number of Students

Table 1

	Minimum	Mean	Maximum
Number of Students	8	55.37802	581

There is an average of 55 students in each course. Least number of students enrolled for a course is 8 and maximum is 581.

- b. Course size – Course Evaluation Score

Table 2

Course Size	18 or less	19 to 28	29 to 60	61 or more
Number of courses in group	112	114	116	113
Minimum course evaluation score	2.3	2.7	2.1	2.8
Mean course evaluation score	4.154464286	4.000877	3.937069	3.9053097
Maximum course evaluation score	5	5	5	4.8

Number of courses and mean course evaluation in each group is similar with little variation. People in 29-60 have given the least course evaluation score of 2.1 and no one with age 61 above has given a full score with the max score being 4.8.

c. Division

Table 3

	No. courses in group	Minimum	Mean	Maximum
Upper Division	300	2.1	3.950667	5
Lower Division	155	2.5	4.091613	5

Majority of the courses are aimed at upper majority and observed to have the least minimum and less mean course score than lower division

d. Gender – Course Evaluation Score

Table 4

	No. courses in group	Minimum	Mean	Maximum
Female	193	2.3	3.897409	4.9
Male	262	2.1	4.073282	5

More than half (57%) of the enrolled students are male and have a greater average and maximum score than females.

e. Gender – Beauty

Table 5

	No. Academics in group	Minimum	Mean	Maximum
Female	39	-1.53884	0.124898	1.8816743
Male	51	-1.51127	-0.11483	1.6859847

Number of male academics is greater while the mean and maximum beauty score for female academics is slightly greater than males

f. Tenure Track

Table 6

	No. Academics in group	Minimum	Mean	Maximum
Tenure Track	75	2.1	3.959773	5
Not TenureTrack	15	2.8	4.133333	5

Vast majority of academics (83%) are on the tenure track but have the least evaluation score as well as mean score when compared to the academics not on tenure track.

g. Education Background

Table 7

	No. Academics in group	Minimum	Mean	Maximum
English education	83	2.1	4.01896	5
Non-English education	7	2.7	3.68928	4.6

Nearly all the academics (barring 7 out of 90) have an English background and have a higher evaluation score as well the maximum score as compared to academics with non-English background

h. Interactions between Tenure Track, Gender and Education Background

Table 8

Tenure Track	Gender	Education	No. Academics in group	Mean
Tenure track	Female	English	28	3.928244
Tenure track	Female	Non English	3	3.716667
Tenure track	Male	English	40	4.020103
Tenure track	Male	Non English	4	3.66875
Not tenure track	Female	English	8	3.86
Not tenure track	Female	Non English	0	0
Not tenure track	Male	English	7	4.396154
Not tenure track	Male	Non English	0	0

Nearly half the academics (44%) can be grouped into having a tenure track with an English background and are males. This group is just second

i. Correlation Analysis

Table 9

Variables	Correlation Coefficient	Two-sided significance
Course evaluation score & course size	-0.172864735	0.000211293
Staff age & beauty	-0.3309946633	0.001439932
Staff age and mean course evaluation score	-0.008123249	0.963060658
Staff beauty and mean course evaluation score	0.178511745	0.09229987

- For the first two comparisons, the two sided significance values are less than 0.05 which indicates a strong statistical significance. However, the correlation coefficient for those comparisons are close to zero (0.172) which means that there isn't a strong relationship between the two values.
- For the next two comparisons, neither are they statistically significant nor they have any strong relation.

3. Discussion and Conclusion

a. Analysis Summary

Majority of the courses are part of the upper division and the teaching staff consists of professors who have completed their undergraduate in an English-speaking country most of whom are male and on tenure track. The ratio of males is greater for both students and professors. In Table 2, it is observed that as the number of students increase, the mean course evaluation decreases. However this inverse relation is not supported by the correlation coefficient which comes out to be -0.172 as

indicated in Table 9. In table 8, lowest evaluation score is observed for male teachers with a non English background followed closely by female teachers in non English background. The mean course evaluation score for all the cases is above 3.6 (lowest being 3.66) which indicates that none of the professors or teachers were rated too harshly.

b. Limitations

From the correlation analysis (Table 9), one can observe that in the given data, none of the factors such as staff age and beauty are statistically significant which indicates that for the given data set, one cannot draw a proper conclusion and we require more data to analyse whether external factors affect the course evaluation score.

References

- [1] Investopedia. 2020. Correlation Coefficient Definition. [online] Available at: <<https://www.investopedia.com/terms/c/correlationcoefficient.asp>> [Accessed 6 June 2020].
- [2] Docs.oracle.com. 2020. Example: Loading Data Into A Table. [online] Available at: <<https://docs.oracle.com/database/121/ADMQS/GUID-7068681A-DC4C-4E09-AC95-6A5590203818.htm#ADMQS0826>> [Accessed 6 June 2020].
- [3] Thatjeffsmith.com. 2020. How To Import From Excel To Oracle With SQL Developer – Thatjeffsmith. [online] Available at: <<https://www.thatjeffsmith.com/archive/2012/04/how-to-import-from-excel-to-oracle-with-sql-developer/comment-page-5/>> [Accessed 6 June 2020].
- [4] Docs.oracle.com. 2020. CORR_*. [online] Available at: <https://docs.oracle.com/cd/B28359_01/server.111/b28286/functions029.htm#SQLRF06314> [Accessed 6 June 2020].

Appendix

SQL Queries –

- Create table queries for importing the data

--Create Table queries

```
CREATE TABLE staff
(
    staffid      NUMBER(5),
    age          NUMBER(5) NOT NULL,
    gender       VARCHAR(1) NOT NULL,
    tenuretrack  NUMBER(1, 0) NOT NULL, --allows storing of binary data
    nonenglish   NUMBER(1, 0) NOT NULL,
    beauty       NUMBER(12, 10) NOT NULL,
    --two digits before the decimal and 10 digits after decimal
    CONSTRAINT staff_pk PRIMARY KEY (staffid)
)

CREATE TABLE course
(
    courseid     NUMBER(5),
    staffid      NUMBER(5),
    courseeval   FLOAT(5) NOT NULL,
    studentcount NUMBER(5) NOT NULL,
    division     VARCHAR(1) NOT NULL,
    CONSTRAINT course_pk PRIMARY KEY (courseid),
    CONSTRAINT fk_staff FOREIGN KEY (staffid) REFERENCES staff(staffid)
)
```

- SQL Analysis Queries

1. Calculate the minimum, mean and maximum number of students in a course.

```
SELECT Min(studentcount),
       Avg(studentcount),
       Max(studentcount)
FROM   course
```

2. Analyse the minimum, mean, and maximum course evaluation score for groups of courses, binned into size groups of 18 or less, 19—28, 29—60, 61 or more.

```
--
fetch number of courses, min, max and avg evaluation score for courses w
ith student count less than or equal to 18
SELECT Count(courseid),
       Min(courseeval),
       Avg(courseeval),
       Max(courseeval)
FROM   course
WHERE  studentcount <= 18
UNION ALL
--
fetch number of courses, min, max and avg evaluation score for courses w
ith student count between 19 to 28
SELECT Count(courseid),
       Min(courseeval),
       Avg(courseeval),
       Max(courseeval)
```

```

FROM    course
WHERE   studentcount >= 19
        AND studentcount <= 28
UNION ALL
--
fetch number of courses, min, max and avg evaluation score for courses with student count between 29 to 60
SELECT  Count(courseid),
        Min(courseeval),
        Avg(courseeval),
        Max(courseeval)
FROM    course
WHERE   studentcount >= 29
        AND studentcount <= 60
UNION ALL
--
fetch number of courses, min, max and avg evaluation score for courses with student count more than 61
SELECT  Count(courseid),
        Min(courseeval),
        Avg(courseeval),
        Max(courseeval)
FROM    course
WHERE   studentcount >= 61

```

3. Analyse minimum, mean, and maximum course evaluation score by division (course level).

--fetch required values for upper division

```

SELECT  division,
        Count(courseid),
        Min(courseeval),
        Avg(courseeval),
        Max(courseeval)
FROM    course
WHERE   division = 'U'
GROUP  BY division
UNION ALL
--fetch required values for lower division
SELECT  division,
        Count(courseid),
        Min(courseeval),
        Avg(courseeval),
        Max(courseeval)
FROM    course
WHERE   division = 'L'
GROUP  BY division

```

4. Analyse minimum, mean and maximum course evaluation score by gender

--fetch required values for femake staff

```

SELECT  gender,
        Count(courseid),
        Min(courseeval),
        Avg(courseeval),
        Max(courseeval)
FROM    course
        inner join staff
            ON course.staffid = staff.staffid
WHERE   gender = 'f'
GROUP  BY gender
UNION ALL

```

```

--fetch requiried values for male staff
SELECT gender,
       Count(courseid),
       Min(courseeval),
       Avg(courseeval),
       Max(courseeval)
FROM   course
       inner join staff
           ON course.staffid = staff.staffid
WHERE  gender = 'm'
GROUP BY gender

```

5. Analyse minimum, mean and maximum beauty by gender.

```

--fetch required details for female staff
SELECT gender,
       Count(staffid),
       Min(beauty),
       Avg(beauty),
       Max(beauty)
FROM   staff
WHERE  gender = 'f'
GROUP BY gender
UNION ALL
--fetch required details for male staff
SELECT gender,
       Count(staffid),
       Min(beauty),
       Avg(beauty),
       Max(beauty)
FROM   staff
WHERE  gender = 'm'
GROUP BY gender

```

6. Analyse minimum, mean and maximum course evaluation by tenure track status.

```

--
distinct staff ID was done as there is repetation of staffID in the cou
rse table
--fetch details for professors on tenure track
SELECT tenuretrack,
       Count(DISTINCT staff.staffid),
       Min(courseeval),
       Avg(courseeval),
       Max(courseeval)
FROM   course
       inner join staff
           ON course.staffid = staff.staffid
WHERE  tenuretrack = 1
GROUP BY tenuretrack
UNION ALL
--fetch details for professors not on tenure track
SELECT tenuretrack,
       Count(DISTINCT staff.staffid),
       Min(courseeval),
       Avg(courseeval),
       Max(courseeval)
FROM   course
       inner join staff
           ON course.staffid = staff.staffid

```

```
WHERE tenuretrack = 0
GROUP BY tenuretrack
```

7. Analyse minimum, mean and maximum course evaluation by education background.

```
--staff with non english undergrad
SELECT nonenglish,
       Count(DISTINCT staff.staffid),
       Min(courseeval),
       Avg(courseeval),
       Max(courseeval)
FROM   course
       inner join staff
           ON course.staffid = staff.staffid
WHERE  nonenglish = 1
GROUP BY nonenglish
UNION ALL
--staff with english undergrad
SELECT nonenglish,
       Count(DISTINCT staff.staffid),
       Min(courseeval),
       Avg(courseeval),
       Max(courseeval)
FROM   course
       inner join staff
           ON course.staffid = staff.staffid
WHERE  nonenglish = 0
GROUP BY nonenglish
```

8. Analyse course evaluation by gender, tenure track, and education background

```
--female staff on tenure track and english undergrad
SELECT tenuretrack,
       gender,
       nonenglish,
       Count(DISTINCT staff.staffid),
       Avg(courseeval)
FROM   course
       inner join staff
           ON course.staffid = staff.staffid
WHERE  tenuretrack = 1
       AND gender = 'f'
       AND nonenglish = '0'
GROUP BY tenuretrack,
       gender,
       nonenglish
UNION ALL
--female staff on tenure track and non english undergrad
SELECT tenuretrack,
       gender,
       nonenglish,
       Count(DISTINCT staff.staffid),
       Avg(courseeval)
FROM   course
       inner join staff
           ON course.staffid = staff.staffid
WHERE  tenuretrack = 1
       AND gender = 'f'
       AND nonenglish = '1'
GROUP BY tenuretrack,
       gender,
```



```

        nonenglish
UNION ALL
--male staff on tenure track and english undergrad
SELECT tenuretrack,
       gender,
       nonenglish,
       Count(DISTINCT staff.staffid),
       Avg(courseeval)
FROM   course
       inner join staff
           ON course.staffid = staff.staffid
WHERE  tenuretrack = 1
       AND gender = 'm'
       AND nonenglish = '0'
GROUP BY tenuretrack,
         gender,
         nonenglish
UNION ALL
--male staff on tenure track and non english undergrad
SELECT tenuretrack,
       gender,
       nonenglish,
       Count(DISTINCT staff.staffid),
       Avg(courseeval)
FROM   course
       inner join staff
           ON course.staffid = staff.staffid
WHERE  tenuretrack = 1
       AND gender = 'm'
       AND nonenglish = '1'
GROUP BY tenuretrack,
         gender,
         nonenglish
UNION ALL
--female staff not on tenure track and english undergrad
SELECT tenuretrack,
       gender,
       nonenglish,
       Count(DISTINCT staff.staffid),
       Avg(courseeval)
FROM   course
       inner join staff
           ON course.staffid = staff.staffid
WHERE  tenuretrack = 0
       AND gender = 'f'
       AND nonenglish = '0'
GROUP BY tenuretrack,
         gender,
         nonenglish
UNION ALL
--female staff not on tenure track and non english undergrad
SELECT tenuretrack,
       gender,
       nonenglish,
       Count(DISTINCT staff.staffid),
       Avg(courseeval)
FROM   course

```

```

        inner join staff
            ON course.staffid = staff.staffid
WHERE tenuretrack = 0
    AND gender = 'f'
    AND nonenglish = '1'
GROUP BY tenuretrack,
        gender,
        nonenglish
UNION ALL
--male staff not on tenure track and english undergrad
SELECT tenuretrack,
        gender,
        nonenglish,
        Count(DISTINCT staff.staffid),
        Avg(courseeval)
FROM course
    inner join staff
        ON course.staffid = staff.staffid
WHERE tenuretrack = 0
    AND gender = 'm'
    AND nonenglish = '0'
GROUP BY tenuretrack,
        gender,
        nonenglish
UNION ALL
--male staff not on tenure track and non english undergrad
SELECT tenuretrack,
        gender,
        nonenglish,
        Count(DISTINCT staff.staffid),
        Avg(courseeval)
FROM course
    inner join staff
        ON course.staffid = staff.staffid
WHERE tenuretrack = 0
    AND gender = 'm'
    AND nonenglish = '1'
GROUP BY tenuretrack,
        gender,
        nonenglish

```

9. calculate the Spearman rank correlation between the four pairs of variables, as follows:

a. Course evaluation score and course size

```

SELECT Corr_s(courseeval, studentcount, 'COEFFICIENT'),
        Corr_s(courseeval, studentcount, 'TWO_SIDED_SIG')
FROM course

```

b. Staff age and beauty

```

SELECT Corr_s(age, beauty, 'COEFFICIENT'),
        Corr_s(age, beauty, 'TWO_SIDED_SIG')
FROM staff

```

c. Staff age and mean course evaluation score

Step 1 – First make a view that stores evaluation score grouped by staff age

```

CREATE VIEW agecorr
(age, avg_score)
AS
SELECT age,
        Avg(courseeval)

```

```

FROM    course
        inner join staff
              ON course.staffid = staff.staffid
GROUP BY age
ORDER BY age

```

Step 2 – Calculate the coefficient using the view

```

SELECT Corr_s(age, avg_score, 'COEFFICIENT'),
       Corr_s(age, avg_score, 'TWO_SIDED_SIG')
FROM   agecorr

```

d. Staff beauty and mean course evaluation score

Step 1 – First make a view that stores evaluation score grouped by staff beauty score

```

CREATE VIEW beautycorr
(beauty, avg_score)
AS
SELECT beauty,
       Avg(courseeval)
FROM   course
        inner join staff
              ON course.staffid = staff.staffid
GROUP BY beauty
ORDER BY beauty

```

Step 2 – Calculate the coefficient using the view

```

SELECT Corr_s(beauty, avg_score, 'COEFFICIENT'),
       Corr_s(beauty, avg_score, 'TWO_SIDED_SIG')
FROM   beautycorr

```