

## ▪ **Data Preparation**

### **1. Introduction**

- ◆ Analysis needs to be undertaken to understand whether particular variables such as class size and the perceived attractiveness of teaching staff influence course evaluations.
- ◆ This investigation will be carried out on the data collected by Daniel Hamermesh and Amy Parker, two researchers from the USA at the University of Texas at Austin, USA, which includes information about 455 courses, taught by teaching staff in various departments. Each course was evaluated on a 5-level ordinal scale with a minimum score of (1) “very unsatisfactory” and a maximum score of (5) “excellent”. Whereas, the attractiveness was rated from 1 (lowest) to 10 (highest) rating scale. The ratings subsequently referred to as “beauty” scores were then normalized to have a mean score of zero. All this information is stored in one file i.e. profEvaluations.csv.

### **2. Relation Schema**

**Professors** (staffid, age, gender, tenuretrack, nonenglish, beauty)

**Course** (id, staffid\*, courseevaluation, students, division)

### **3. Importing data into tables**

#### **Professors table**

- Once the connection is successfully established, make sure that the profEvaluations.csv file is in appropriate directory to begin with. Navigate to the connection in the left-pane and right click on the connection, then click on ‘import data’. The open dialog box appears.
- Browse and select the profEvaluations.csv file and then click Open. The Data Import Wizard appears, with the Data Preview page displayed. Make sure that Header is deselected, Format is set to csv and that Left Enclosure and Right Enclosure is None. Then click Next.
- The import page appears, enter the table name i.e. ‘Professors’ and let the import method be insert. Then click Next.
- In Choose column page, select the columns corresponding to the ‘Professors’ relation schema and click Next. In Column Definition page, iterate through each column and alter the datatype and the size appropriately depending on the schema and its use.
- Click Finish, the new ‘Professors’ table will be created by successfully importing the data.
- Once the table is imported, first run some queries to check if there are any duplicate records. If duplicate records are found then remove them.
- Now set the primary key by referring to the ‘Professors’ relation schema.

#### **Course table**

- Perform all the above mentioned steps for creating the ‘Course’ table but select the columns corresponding to the ‘Course’ schema. Also, choose appropriate datatype and size for each column while iterating through the columns.
- After that click Finish, the new the new ‘Course’ table will be created by successfully importing the data.
- Check for any duplicates in course table as well. If duplicates records are found then eliminate them. Then set the primary key and foreign key depending on the ‘Course’ relation schema.

▪ **Analysis**

- **Course Sizes – Number of Students:** The table contains the minimum, mean and maximum number of students in the course.

	<b>Minimum</b>	<b>Mean</b>	<b>Maximum</b>
<b>Number of students</b>	8	55.3780	581

- **Course Sizes – Course Evaluation Score:** The table represents the minimum, mean and maximum course evaluation score given by students which are then grouped by the age of the staff members. We can see that, as the age group increases, the mean evaluation course decreases

<b>Course size</b>	<b>18 or less</b>	<b>19-28</b>	<b>29-60</b>	<b>61 or more</b>
<b>Number of courses in group</b>	112	114	116	113
<b>Minimum course evaluation score</b>	2.3	2.7	2.1	2.8
<b>Mean course evaluation score</b>	4.15	4.00	3.93	3.90
<b>Maximum course evaluation score</b>	5	5	5	4.8

- **Division:** The table represents the minimum, mean and maximum course evaluation score which are then grouped by the division of students. We can see that there are more courses in group in upper division (300) compare to lower division (155). However, the courses in lower division have better evaluation score with minimum as 2.5 and mean as 4.09.

	<b>No. courses in group</b>	<b>Minimum</b>	<b>Mean</b>	<b>Maximum</b>
<b>Upper division</b>	300	2.1	3.95	5
<b>Lower division</b>	155	2.5	4.09	5

- **Gender – Course Evaluation Score:** The table represents the minimum, mean and maximum course evaluation score grouped by gender of the staff. We can see that, based on the mean, Male staff (mean as 4.073) tends to have a better evaluation score as compared to Female staff (mean as 3.897).

	<b>No. courses in group</b>	<b>Minimum</b>	<b>Mean</b>	<b>Maximum</b>
<b>Female</b>	193	2.3	3.897	4.9
<b>Male</b>	262	2.1	4.073	5

- **Gender – Beauty:** The table represents the minimum, mean and maximum beauty i.e. attractiveness score of staff grouped by gender of the staff. We can see that female staff on an average has a better beauty score of 0.125 compared to Male staff.

	<b>No. academics in group</b>	<b>Minimum</b>	<b>Mean</b>	<b>Maximum</b>
--	-------------------------------	----------------	-------------	----------------

<b>Female</b>	39	-1.539	0.125	1.882
<b>Male</b>	51	-1.511	-0.115	1.686

- **Tenure track:** The table shows the minimum, mean and maximum course evaluation score grouped by tenure of the staff. The table illustrates that the staff members who don't have tenure had a better course evaluation score with minimum score of 2.8 and mean score of 4.13.

	<b>No. academics in group</b>	<b>Minimum</b>	<b>Mean</b>	<b>Maximum</b>
<b>Tenure track</b>	75	2.1	3.96	5
<b>Not Tenure track</b>	15	2.8	4.13	5

- **Education Background:** The given table displays the minimum, mean and maximum course evaluation score grouped by the completion of staff's undergraduate education in an English or a non-English speaking country. As we can see, the staff who graduated in an English speaking country had score well in their course evaluation score with a mean and max score of 4.012 and 5 respectively. Whereas, the Non-English speaking staff didn't score well with mean and max score as 3.689 and 4.6 respectively.

	<b>No. academics in group</b>	<b>Minimum</b>	<b>Mean</b>	<b>Maximum</b>
<b>English education</b>	83	2.1	4.012	5
<b>Non - English education</b>	7	2.7	3.689	4.6

- **Interactions between Tenure Track, Gender and Education Background:** The table represents the, gender, tenure and education background of staff with their course evaluation score. The male staff with no tenure and from English speaking background has the highest mean course evaluation score of 4.396.

<b>Tenure track</b>	<b>Gender</b>	<b>Education</b>	<b>No. academics in group</b>	<b>Mean</b>
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
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:** The following table shows the result of correlation analysis between different variable. There is a negative relation between ‘Course evaluation score and course size’, ‘Staff age and beauty’ and ‘Staff age and mean course evaluation score’ which is -0.17286, -0.33099 and -0.02214 respectively. Also, the ‘Staff beauty and mean course evaluation score’ have a positive relation i.e. 0.17851. However, the two sided significance for ‘Course evaluation score and course size’ and ‘Staff age and beauty’ is  $< 0.05$ . Therefore, the correlation isn’t statistically 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 statistically 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	-0.02214	0.83594
Staff beauty and mean course evaluation score	0.17851	0.0923

## ▪ Discussions and Conclusions

- There is a positive relation between staff attractiveness and mean course evaluation score; this means that the staff attractiveness does influence the course evaluations. A negative relation exists between staff age and mean course evaluation score, beauty; this means that younger the staff better course evaluation score and more attractiveness.
- The male staff with no tenure and from English speaking background had the highest mean course evaluation score of 4.396. This means that staff from non-English background and with no tenure had better performance based on the evaluation from the students.
- Correlation research suggests that there is a relationship between two variables but it cannot prove that one variable causes a change in another variable (i.e. causation). Correlation only predicts relationship rather not cause and effect. Also, Correlation doesn’t allow us to go beyond the data that is given.

## ▪ References

- 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.
- McLeod, S. A. (2018, January 14). Correlation definitions, examples & interpretation. Simply Psychology. Available at <https://www.simplypsychology.org/correlation.html> > [Accessed 9 2020]

▪ **Appendix**

Serial No	Question	SQL Queries
1	Check 'Professors' table for duplicate records	Select staffid, count(*) from professors Having count(*) > 1 Group by staffid;
2	Removing duplicate records from 'Professors' table	Delete professors p1 Where rowid > (Select Min(rowid) From professors p2 Where p2.staffid = p1.staffid);
3	Setting 'staffid' as primary key for 'Professors' table	Alter Table professors Add Constraint staff_p primary key (staffid);
4	Check 'Course' table for duplicate records	Select id, count(*) From course Having count(*) > 1 Group by id;
5	Setting primary and foreign key for 'Course' table	Alter Table course Add Constraint id_p primary key (id);  Alter Table course Add Constraint staff_f foreign key (staffid) References professors(staffid);
6	Course Sizes – Number of Students	Select Max(students), Avg(students), Min(students) From course;
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

		<p>Where students between 29 and 60;</p> <p>Select Count(students), Max(courseevaluation), Avg(courseevaluation), Min(courseevaluation) From course Where students &gt; 60;</p>
8	Division	<p>Select Count(courseevaluation), Max(courseevaluation), Avg(courseevaluation), Min(courseevaluation) From course Where Upper(division) = 'U';</p> <p>Select Count(courseevaluation), Max(courseevaluation), Avg(courseevaluation), Min(courseevaluation) From course Where upper(division) = 'L';</p>
9	Gender – Course Evaluation Score	<p>Select Count(c.courseevaluation), Max(c.courseevaluation), Avg(c.courseevaluation), Min(c.courseevaluation) From course c Join professors p on c.staffid = p.staffid Where Upper(p.gender) = 'F';</p> <p>Select Count(c.courseevaluation), Max(c.courseevaluation), Avg(c.courseevaluation), Min(c.courseevaluation) From course c Join professors p on c.staffid = p.staffid where Upper(p.gender) = 'M';</p>
10	Gender – Beauty	<p>Select Count(p.beauty), Max(p.beauty), Avg(p.beauty), Min(p.beauty) From professors p Where Upper(p.gender) = 'F';</p> <p>Select Count(p.beauty), Max(p.beauty), Avg(p.beauty), Min(p.beauty) From professors p Where Upper(p.gender) = 'M';</p>
11	Tenure track	<p>Select Count(Distinct(c.courseevaluation)), Max(c.courseevaluation), Avg(c.courseevaluation), Min(c.courseevaluation) From course c Join professors p on c.staffid = p.staffid where p.tenuretrack = 1;</p>

		<p>Select Count(Distinct(c.courseevaluation)),  Max(c.courseevaluation),  Avg(c.courseevaluation), Min(c.courseevaluation)  From course c  Join professors p on c.staffid = p.staffid  where p.tenuretrack = 0;</p>
12	Education Background	<p>Select Count(Distinct(c.courseevaluation)),  Max(c.courseevaluation),  Avg(c.courseevaluation), Min(c.courseevaluation)  From course c  Join professors p on c.staffid = p.staffid  where p.nonenglish = 0;</p> <p>Select Count(Distinct(c.courseevaluation)),  Max(c.courseevaluation),  Avg(c.courseevaluation), Min(c.courseevaluation)  From course c  Join professors p on c.staffid = p.staffid  where p.nonenglish = 1;</p>
13	Interactions between Tenure Track, Gender and Education Background	<p>Select Count(Distinct(c.courseevaluation)),  Avg(c.courseevaluation)  From course c  Join professors p on c.staffid = p.staffid  Where p.tenuretrack = 0  And Upper(p.gender) = 'F'  And p.nonenglish = 0 ;</p> <p>Select Count(Distinct(c.courseevaluation)),  Avg(c.courseevaluation)  From course c  Join professors p on c.staffid = p.staffid  Where p.tenuretrack = 0  And Upper(p.gender) = 'F'  And p.nonenglish = 1 ;</p> <p>Select count(Distinct(c.courseevaluation)),  avg(c.courseevaluation)  From course c  Join professors p on c.staffid = p.staffid  Where p.tenuretrack = 0  And Upper(p.gender) = 'M'  And p.nonenglish = 0 ;</p> <p>Select Count(Distinct(c.courseevaluation)),  Avg(c.courseevaluation)  From course c  Join professors p on c.staffid = p.staffid  Where p.tenuretrack = 0  And Upper(p.gender) = 'M'</p>

		<p>And p.nonenglish = 1 ;</p> <p>Select Count(Distinct(c.courseevaluation)),  Avg(c.courseevaluation)  From course c  Join professors p on c.staffid = p.staffid  Where p.tenuretrack = 1  And Upper(p.gender) = 'F'  And p.nonenglish = 0 ;</p> <p>Select Count(Distinct(c.courseevaluation)),  Avg(c.courseevaluation)  From course c  Join professors p on c.staffid = p.staffid  Where p.tenuretrack = 1  And Upper(p.gender) = 'F'  And p.nonenglish = 1 ;</p> <p>Select Count(c.courseevaluation),  Avg(c.courseevaluation)  From course c  Join professors p on c.staffid = p.staffid  Where p.tenuretrack = 1  And Upper(p.gender) = 'M'  And p.nonenglish = 0 ;</p> <p>Select Count(Distinct(c.courseevaluation)),  Avg(c.courseevaluation)  From course c  Join professors p on c.staffid = p.staffid  Where p.tenuretrack = 1  And Upper(p.gender) = 'M'  And p.nonenglish = 1 ;</p>
--	--	---



14	Correlation Analysis	<p>Select Round(Corr_s(courseevaluation, students, 'coefficient'), 5) coefficient,  Round(Corr_s(courseevaluation , students, 'two_sided_sig'), 5) two_sided_p_value from course;</p> <p>Select Round(Corr_s(age, beauty, 'coefficient'), 5) coefficient,  Round(Corr_s(age , beauty, 'two_sided_sig'), 5) two_sided_p_value  From professors;</p> <p>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 professors  Group by age , staffid;</p> <p>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 professors  Group by beauty, staffid;</p>
----	----------------------	---