DMDD ASSIGNMENT 3

TOPIC NAME: University Recommendation System

Github Repository:- https://github.com/Muskansri1/University_Recommendation_System

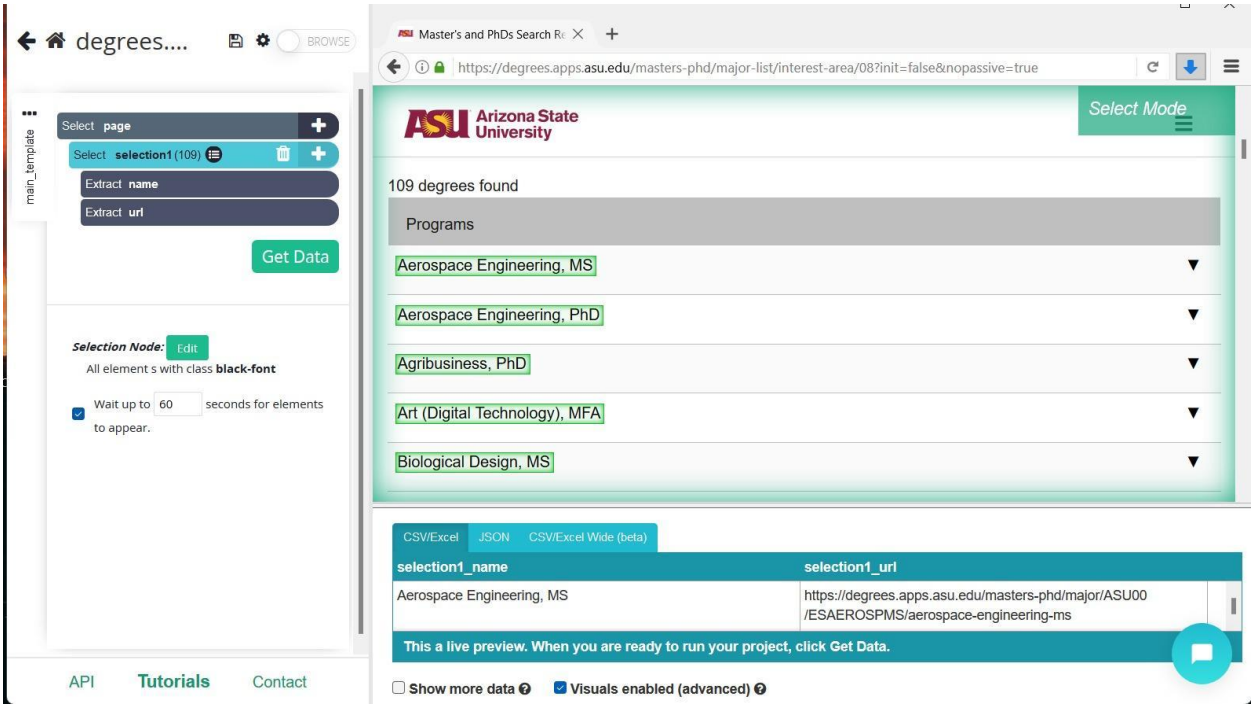
Members:

- Sameer Nimse (002752914)
- Muskan Srivastava (002794929)

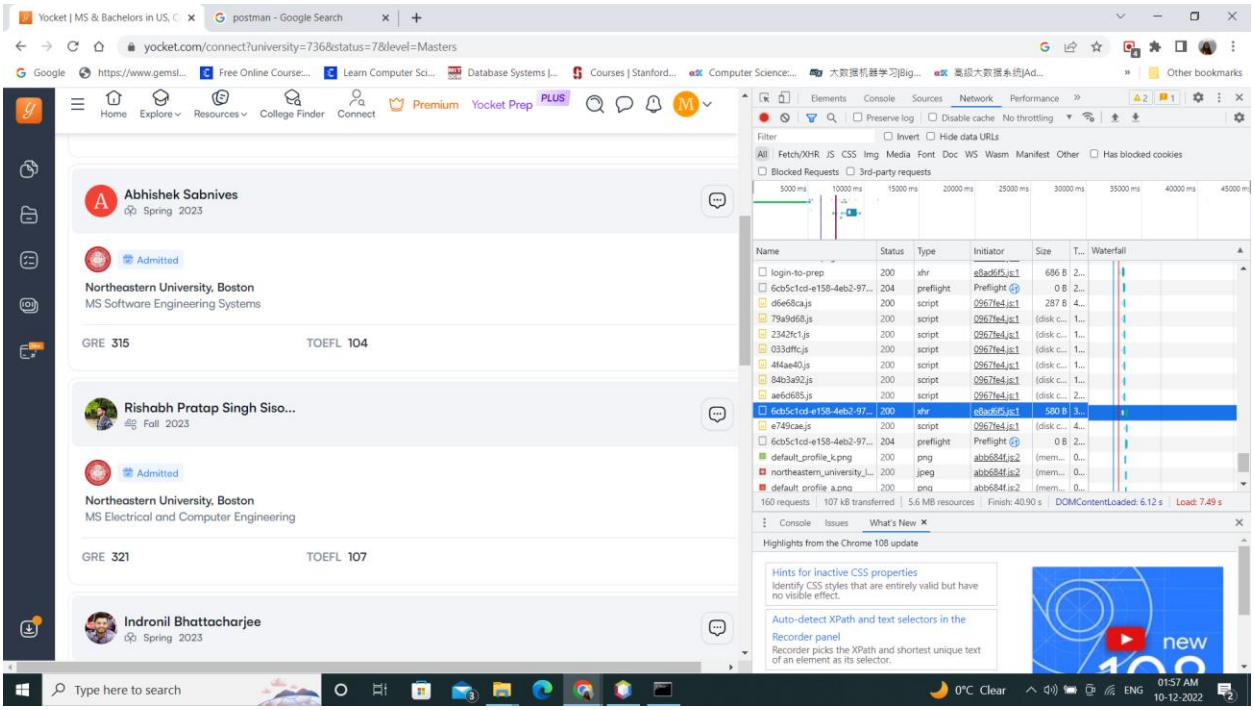
1. Sources of data:

For this assignment, we scrapped the data using Python scripts and bots. We have also used Parsehub, a free web scraping tool to turn any site into a spreadsheet or API. We scraped information about the universities from yocket, QS ranking, wikipedia, edulix, admits.fyi.
We have also scrapped the data using APIs. We fetched the JSON files using Postman. All the data scrapped and fetched are real time and dynamic. We have also imported data from Kaggle which was scrapped from Edulix.com as its original source.

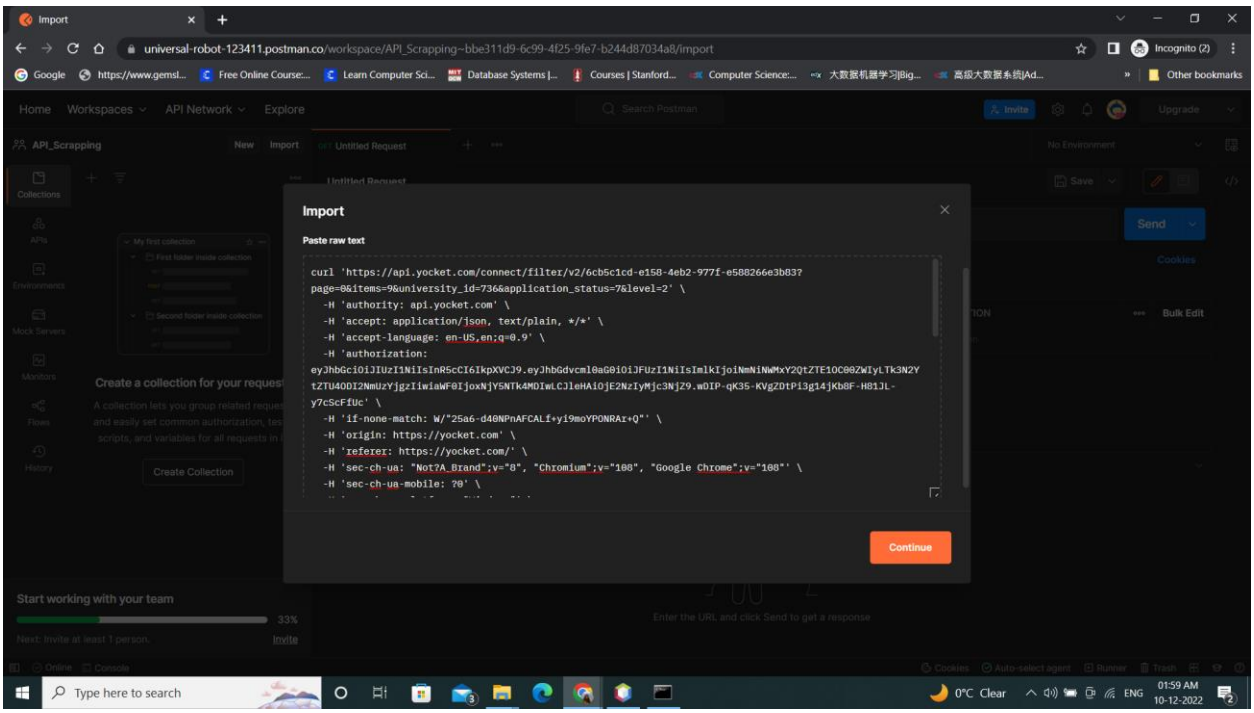
Below is a screenshot of the Parsehub tool:

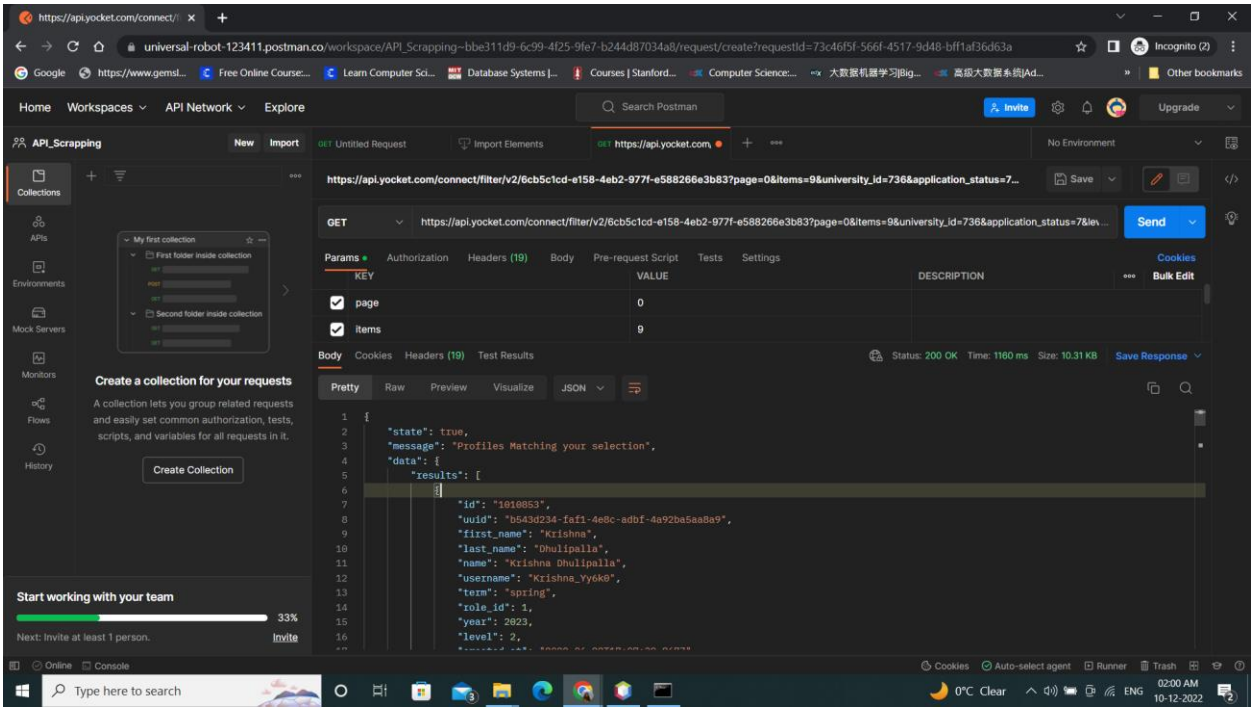


Through APIs:

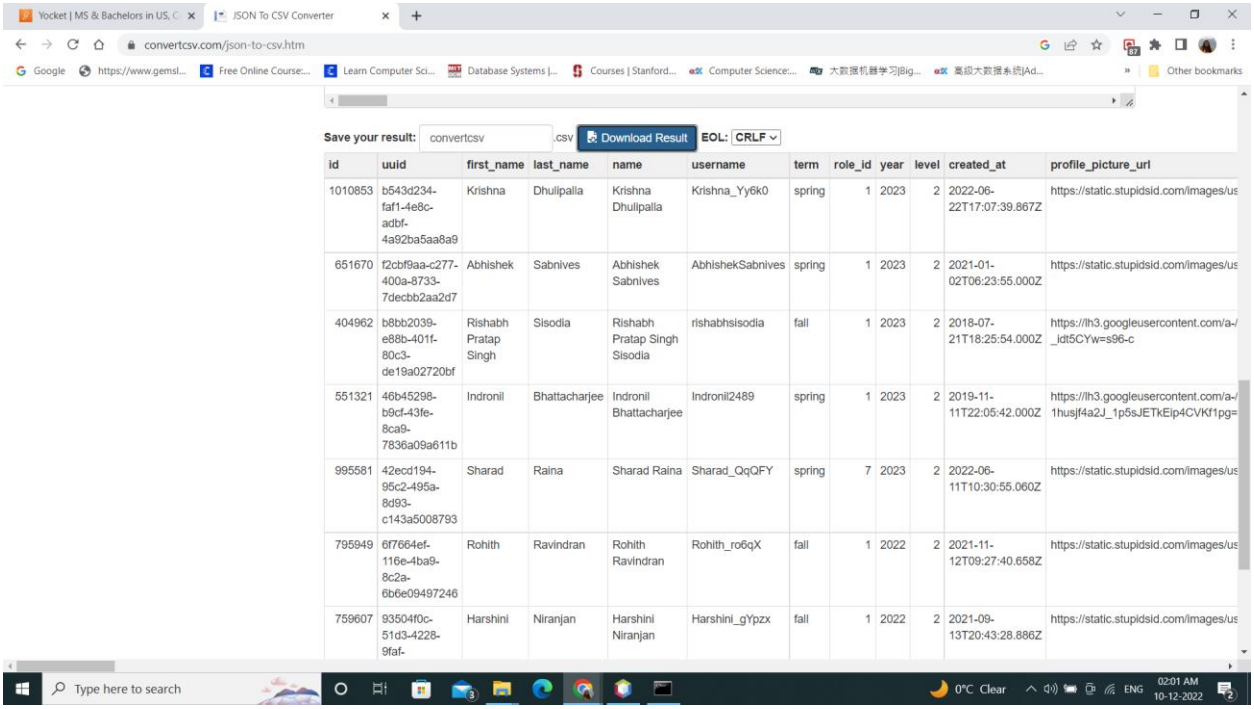


Finding the xhr type files for the API scrapping to import in Postman and send a get request.

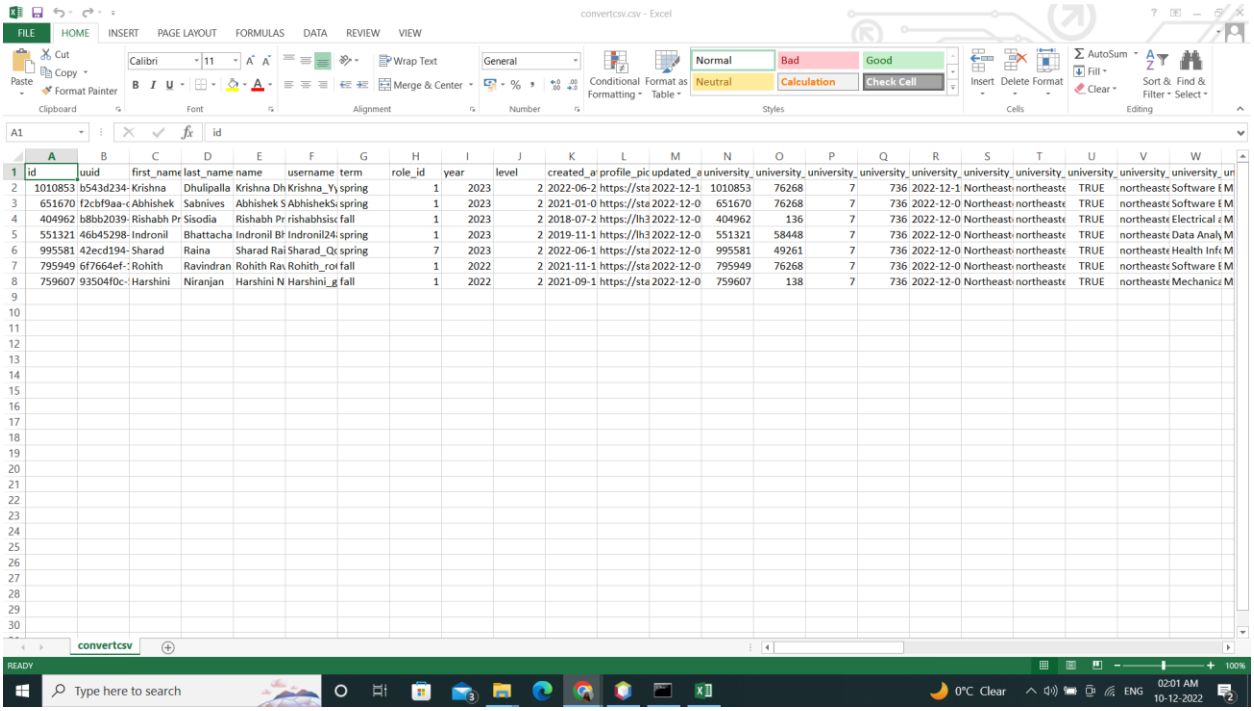




Converting from JSON to CSV.



Result:



id	uuid	first_name	last_name	name	username	term	role_id	year	level	created_a	profile_pic	updated_a	university_	university_	university_	university_	university_	university_	university_	university_	university_	university_	un
1010853	b543d234-	Krishna	Dhulipalla	Krishna Dh	Krishna_Yy	spring	1	2023	2	2022-06-2	https://sta	2022-12-1	1010853	76268	7	736	2022-12-1	Northeast	northeaste	TRUE	northeaste	Software E	M
651670	f2cbf9aa-(	Abhishek	Sabnives	Abhishek S	AbhishekS	spring	1	2023	2	2021-01-0	https://sta	2022-12-0	651670	76268	7	736	2022-12-0	Northeast	northeaste	TRUE	northeaste	Software E	M
404962	b8bb2039-	Rishabh Pr	Sisodia	Rishabh Pr	rishabhsisc	fall	1	2023	2	2018-07-2	https://lh3	2022-12-0	404962	136	7	736	2022-12-0	Northeast	northeaste	TRUE	northeaste	Electrical a	M
551321	46b45298-	Indronil	Bhattacha	Indronil Bh	Indronil24	spring	1	2023	2	2019-11-1	https://lh3	2022-12-0	551321	58448	7	736	2022-12-0	Northeast	northeaste	TRUE	northeaste	Data Analy	M
995581	42ecd194-	Sharad	Raina	Sharad Rai	Sharad_Qc	spring	7	2023	2	2022-06-1	https://sta	2022-12-0	995581	49261	7	736	2022-12-0	Northeast	northeaste	TRUE	northeaste	Health Infc	M
795949	6f7664ef-:	Rohith	Ravindran	Rohith Ra	Rohith_roi	fall	1	2022	2	2021-11-1	https://sta	2022-12-0	795949	76268	7	736	2022-12-0	Northeast	northeaste	TRUE	northeaste	Software E	M
759607	93504f0c-:	Harshini	Niranjan	Harshini N	Harshini_g	fall	1	2022	2	2021-09-1	https://sta	2022-12-0	759607	138	7	736	2022-12-0	Northeast	northeaste	TRUE	northeaste	Mechanica	M

2. Auditing

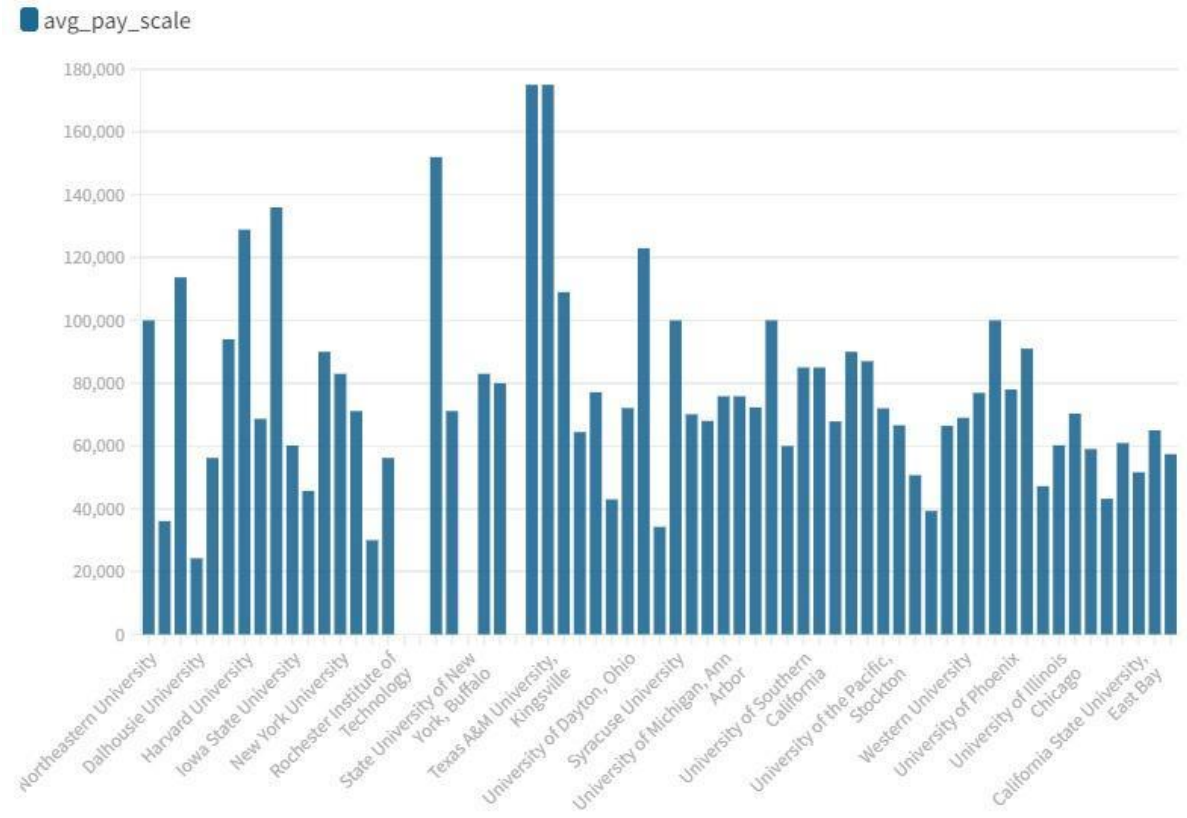
Example 1:

Below is an example of the Average pay scale of a graduate student vs the university from which he graduated. The below value ranges from 80,000 dollars to 130,000 dollars which is the considered to be the average pay scale for Masters student in the engineering field.

Master's in Engineering Salary

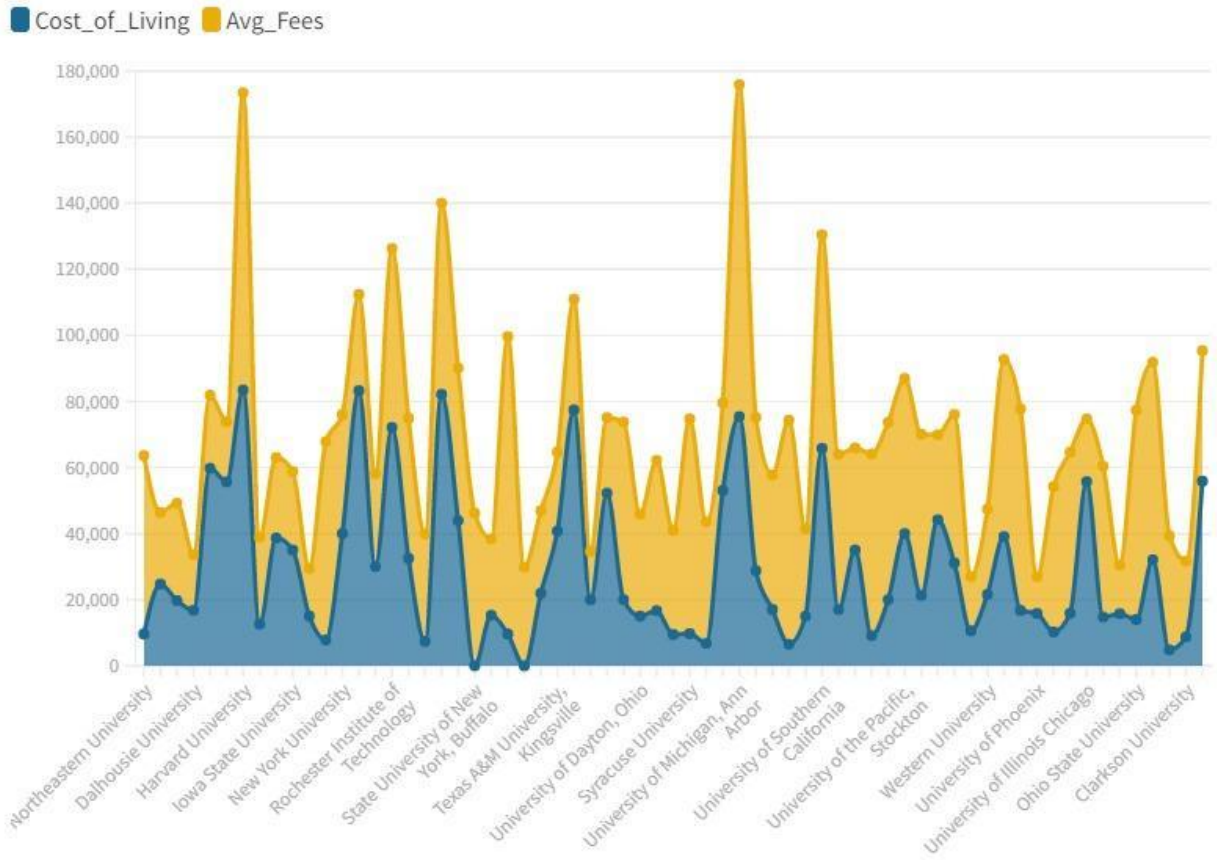
While there are numerous reasons to continue your engineering education, one of the major advantages of completing a master's degree in engineering is the positive impact it can have on your earning potential. Though engineering positions in general are already among some of the highest paid in the country, earning a second degree can help you attain an even more competitive salary in a variety of positions.

According to the salary aggregate site Payscale, the median annual master's in engineering salary was approximately $100,000 as of August 2022. This is a noticeable increase from the median annual salary for a bachelor's in engineering, which Payscale reported at approximately $87,000. While factors such as years of experience and the specific branch of engineering all affect salaries, earning an advanced degree typically provides a financial boost.



Example 2:

Below is an example of cost of living and average fees for a course at a particular university. We can see that the average Fees for the universities in the US is approximately 50,000 to 80,000 dollars whereas the average cost of living in the US is approx 20,000 dollars for the entire course. Hence validating the data fetched.



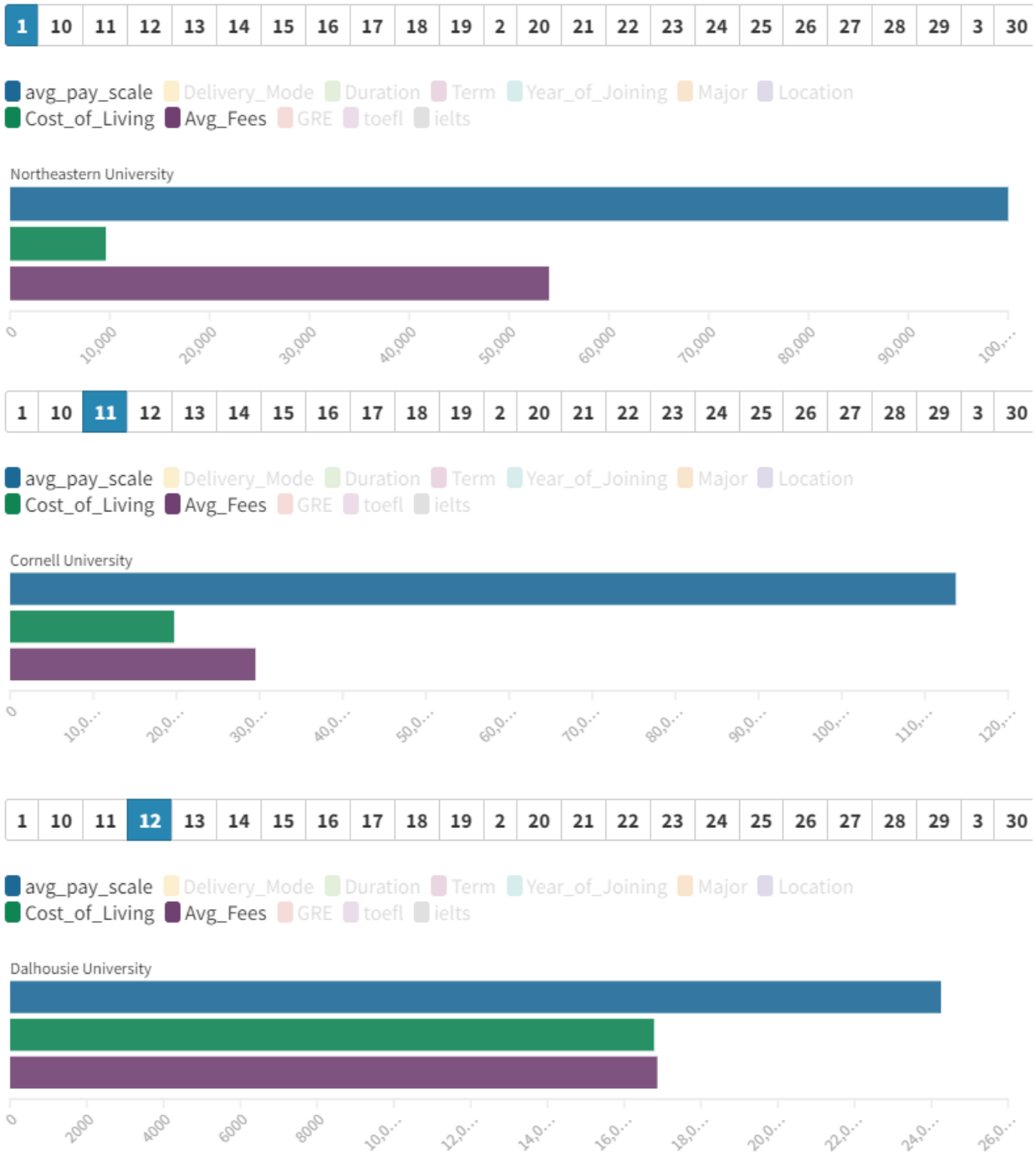
Example 3:

Below are the examples below that reflect the comparison between average pay scale after graduation with the cost of living and average fees of the university. graduation vs the cost of living and the average fees.

Taking Northeastern into consideration:

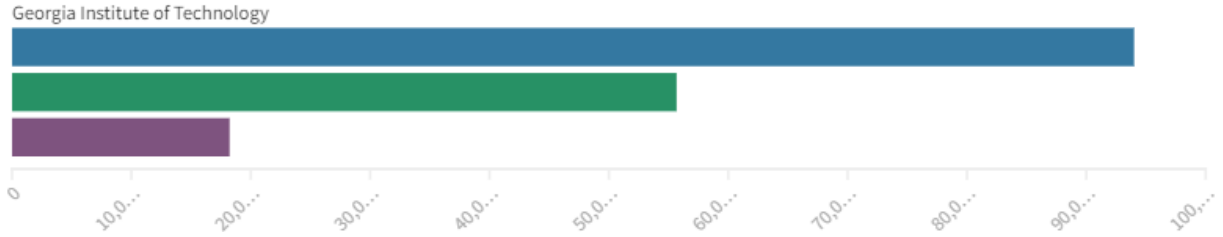
Degree	Average Salary
Doctorate	$120,000
MBA	$112,000
Master (Other)	$90,000
Masters in Science	$81,000
Bachelor of Science	$80,000
Masters in Management	$76,000

Engineering	$1,740 (per credit hour)
Health Informatics	$1,350 (per credit hour)
Khoury College of Computer Sciences	$1,667 (per credit hour)



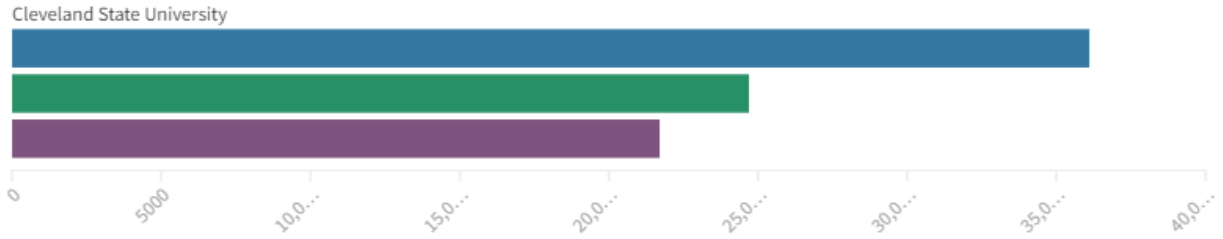
1	10	11	12	13	14	15	16	17	18	19	2	20	21	22	23	24	25	26	27	28	29	3	30

avg_pay_scale Delivery_Mode Duration Term Year_of_Joining Major Location
Cost_of_Living Avg_Fees GRE toefl ielts



1	10	11	12	13	14	15	16	17	18	19	2	20	21	22	23	24	25	26	27	28	29	3	30

avg_pay_scale Delivery_Mode Duration Term Year_of_Joining Major Location
Cost_of_Living Avg_Fees GRE toefl ielts



4. Completeness:

The degree to which every piece of data in a data set is readily accessible is referred to as data completeness in the data quality framework. The percentage of missing data entries serves as a gauge of data completeness. For example, a 100-field column with 30 blank fields has an 70% completeness level. Depending on your industry, 30% of missing data could be potentially harmful and could cause loss of worth hundreds of thousands of dollars.

Data completeness corresponds to getting the data that is important to the user rather than having 100% data.

For example:

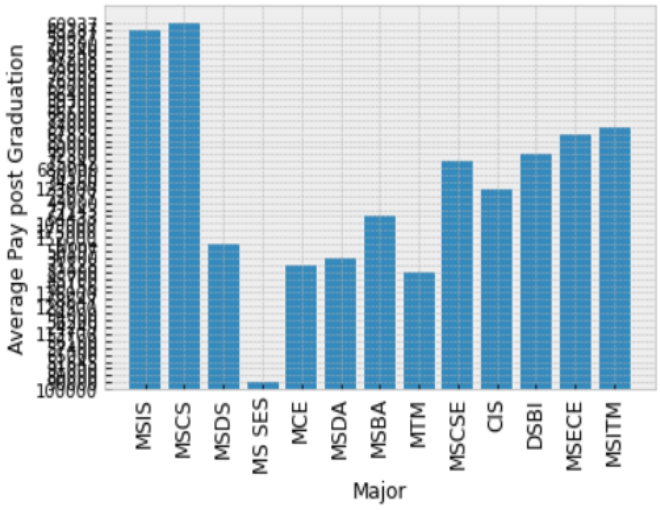
The average pay below displays the wide set of dataset considered from the past 10 years due to which, we can see the trend in data of the courses and data set knowledge availability.

The availability of data for courses like MSIS, MSCS increase the ratio of knowledge gathered with respect to the student's admitted in the last 10 years.

In [15]:
```
xvalue = University_df['Major']
yvalue = University_df['avg_pay_scale']

plt.xlabel('Major')
plt.ylabel('Average Pay post Graduation')
plt.xticks(rotation=90, fontsize = 'large')
plt.bar(xvalue,yvalue)
```

Out[15]:
```
<BarContainer object of 65 artists>
```



4. Data Consistency:

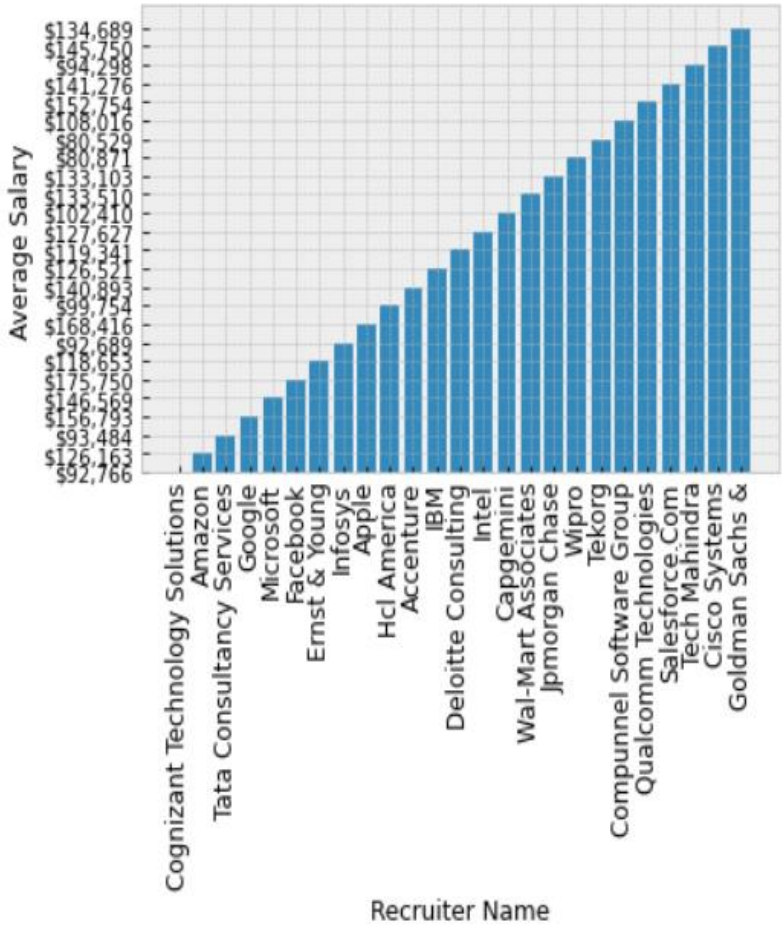
Below is the average salary of the top recruiters and the growth in the average pay scale of the recruiters. The data belows shows a consistent and steady growth which is reflected by the current market scenarios.

The average salary for all the recruiters is reflected in dollars therefore the data is also consistent in the metrics.

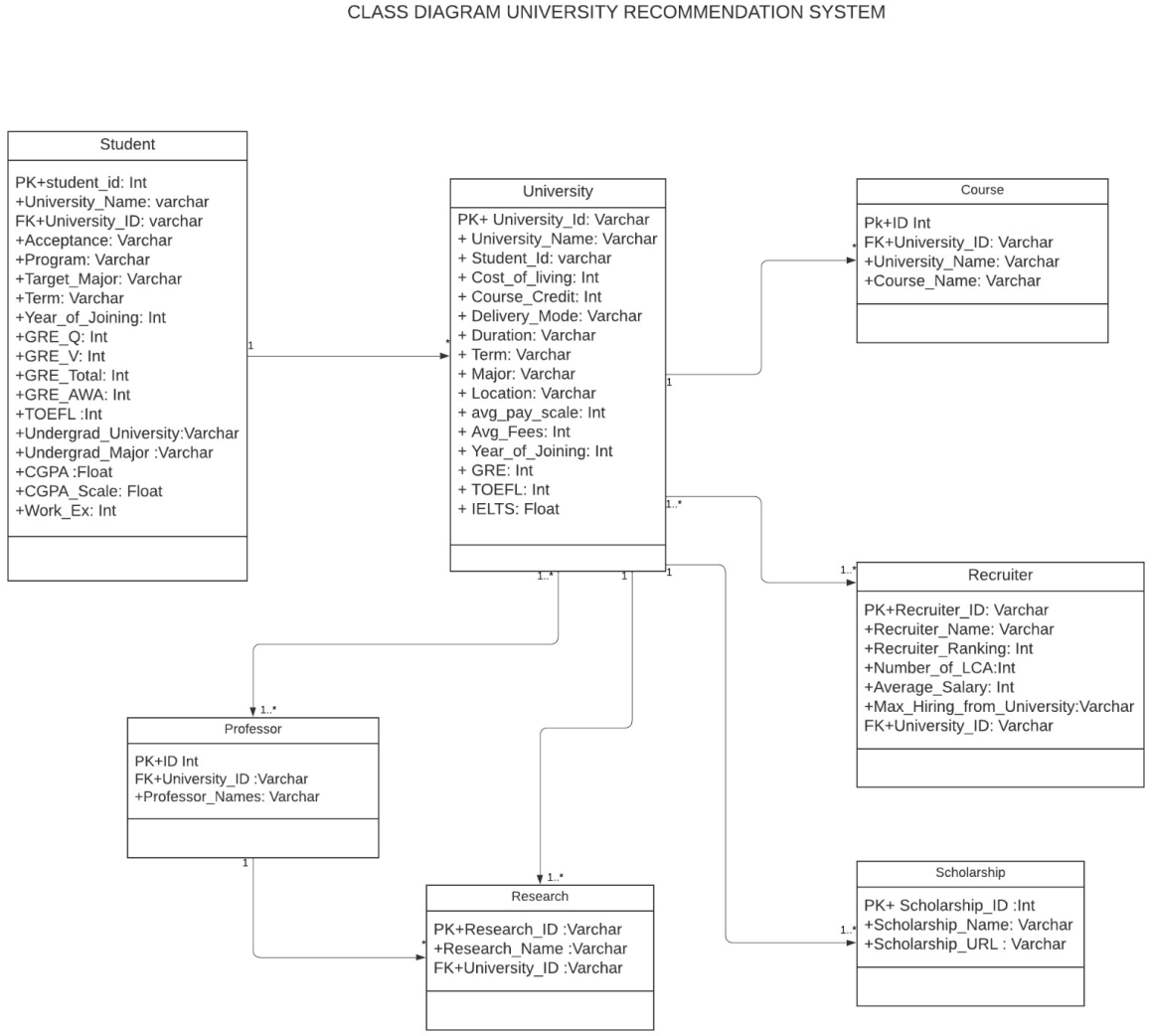
```
xvalue = Recruiter_df['Recruiter_Name']
yvalue = Recruiter_df['Average_Salary']

plt.xlabel('Recruiter Name')
plt.ylabel('Average Salary')
plt.xticks(rotation=90, fontsize = 'large')
plt.bar(xvalue,yvalue)
```

```
<BarContainer object of 25 artists>
```



5. ER Diagram with the feedback from Assignment -2



6 .SQL to create and insert the data into your database

University Table

```
CREATE TABLE university(
University_ID VARCHAR(255),
University_Name VARCHAR(255),
avg_pay_scale INT,
Delivery_Mode VARCHAR(255),
Duration VARCHAR(255),
Term VARCHAR(255),
Year_of_Joining INT,
Major VARCHAR(255),
Location VARCHAR(255),
Cost_of_Living INT,
GRE INT,
TOEFL INT,
IELTS FLOAT,
PRIMARY KEY (University_ID)
);
```

Student Table

```
CREATE TABLE Student (
student_id INT,
University_Name varchar(255),
University_ID INT,
Acceptance VARCHAR(255),
Program VARCHAR(255),
Target_Major VARCHAR(255),
Term VARCHAR(255),
Year_of_Joining INT,
GRE_Q INT,
GRE_V INT,
GRE_Total INT,
GRE_AWA INT,
TOEFL INT,
Undergrad_University VARCHAR(255),
Undergrad_Major VARCHAR(255),
CGPA Float,
```

```
CGPA_Scale Float,
Work_Ex INT,
PRIMARY KEY (student_id),
FOREIGN KEY (University_ID) REFERENCES User(University_ID)
);
```

Scholarship Table

```
CREATE TABLE Scholorship (
Scholarship_ID INT,
Scholarship_Name Varchar(255),
Scholarship_URL Varchar (255),
PRIMARY KEY (Scholarship_ID)
);
```

Research Table

```
CREATE TABLE Research(
Research_ID Varchar(255),
Research_Name VARCHAR(255),
University_ID Varchar(255),
PRIMARY KEY (Research_ID),
FOREIGN KEY (University_ID) REFERENCES university(University_ID)
);
```

Course Table

```
CREATE TABLE Course (
ID INT,
University_ID VARCHAR(255),
University_Name VARCHAR(255),
Course_Name VARCHAR(255),
PRIMARY KEY (ID),
FOREIGN KEY (University_ID) REFERENCES university(University_ID)
);
```

Professor Table

```
CREATE TABLE Professor (
ID INT,
```

```
University_ID VARCHAR(255),
Professor_Names VARCHAR(255),
PRIMARY KEY (ID),
FOREIGN KEY (University_ID) REFERENCES university(University_ID)
);
```

Student Table

```
CREATE TABLE Student (
student_id INT,
University_Name varchar(255),
University_ID varchar(255),
Acceptance VARCHAR(255),
Program VARCHAR(255),
Target_Major VARCHAR(255),
Term VARCHAR(255),
Year_of_Joining INT,
GRE_Q INT,
GRE_V INT,
GRE_Total INT,
GRE_AWA INT,
TOEFL INT,
Undergrad_University VARCHAR(255),
Undergrad_Major VARCHAR(255),
CGPA Float,
CGPA_Scale Float,
Work_Ex INT,
PRIMARY KEY (student_id),
FOREIGN KEY (University_ID) REFERENCES University (University_ID)
);
```

Recruiter Table

```
CREATE TABLE Recruiter(
Recruiter_ID Varchar(255),
Recruiter_Name Varchar(255),
```

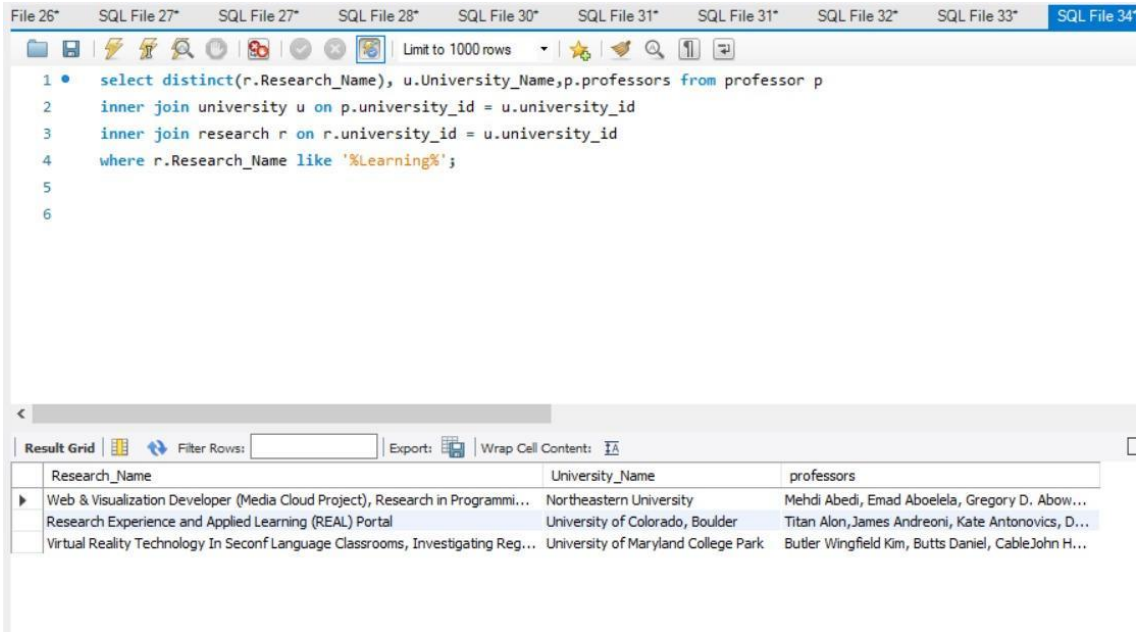
```
Recruiter_Ranking INT,
Number_of_LCA INT,
Average_Salary INT,
Max_Hiring_from_University Varchar(255),
University_ID Varchar(255),
PRIMARY KEY (Recruiter_ID),
FOREIGN KEY (University_ID) REFERENCES university(University_ID)
);
```

7. USE CASES:-

1. **Search for a distinct research opportunity (for example – Machine Learning) available in a particular university along with the list of all the professors of that university.**
   Description: This query returns a list of ongoing research in a university under a professor
   Actor: Student
   Precondition: The student must select research opportunities
   Steps:
   Actor Action: The student views the ongoing research from a university
   System Response: The system displays all the ongoing research from a university under the professor
   Postcondition: The system generates a list of ongoing research

   SQL Query

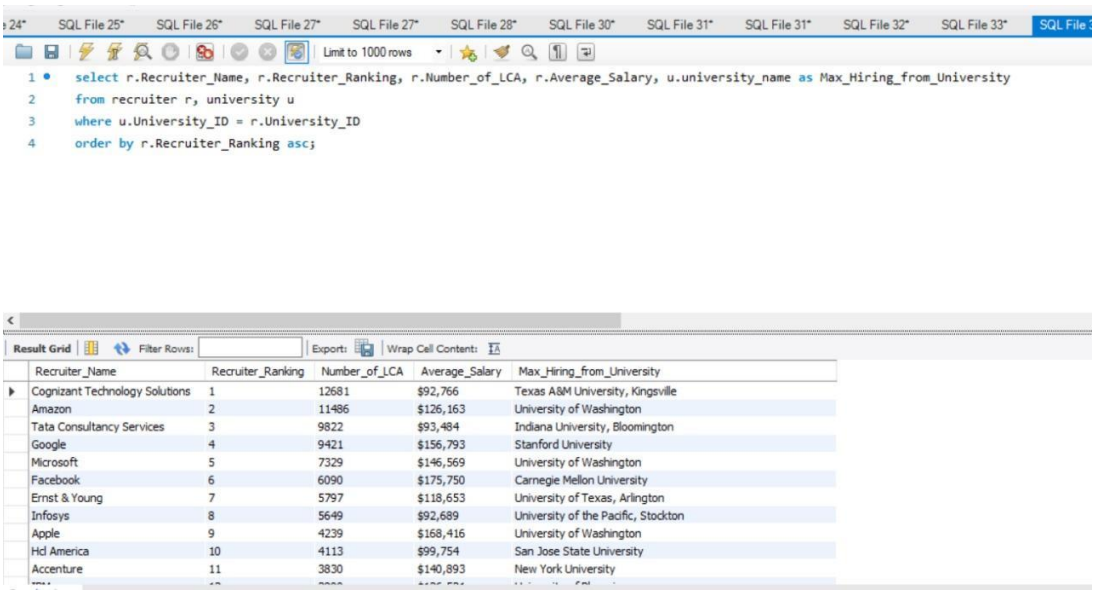
```
select distinct(r.Research_Name), u.University_Name,p.professors from professor p
inner join university u on p.university_id = u.university_id
inner join research r on r.university_id = u.university_id
where r.Research_Name like '%Learning%';
```



2. **Display the recruiter ranking along with the average salary and the university from which it hires the maximum number of students**

   Description: Display a list of recruiters along with the average salary that they offer and the maximum number of students they hire from a particular university
   Actor: Student
   Precondition: The student must select a list of top recruiters
   Steps:
   Actor Action: The student views the recruiters along with the average package
   System Response: The system displays a list of all the recruiters and the university from which they hire the maximum number of students
   Postcondition: The system generates a list of recruiters

   SQL Query:
   select r.Recruiter_Name, r.Recruiter_Ranking, r.Number_of_LCA, r.Average_Salary, u.university_name as Max_Hiring_from_University from recruiter r, university u where u.University_ID = r.University_ID order by r.Recruiter_Ranking asc;

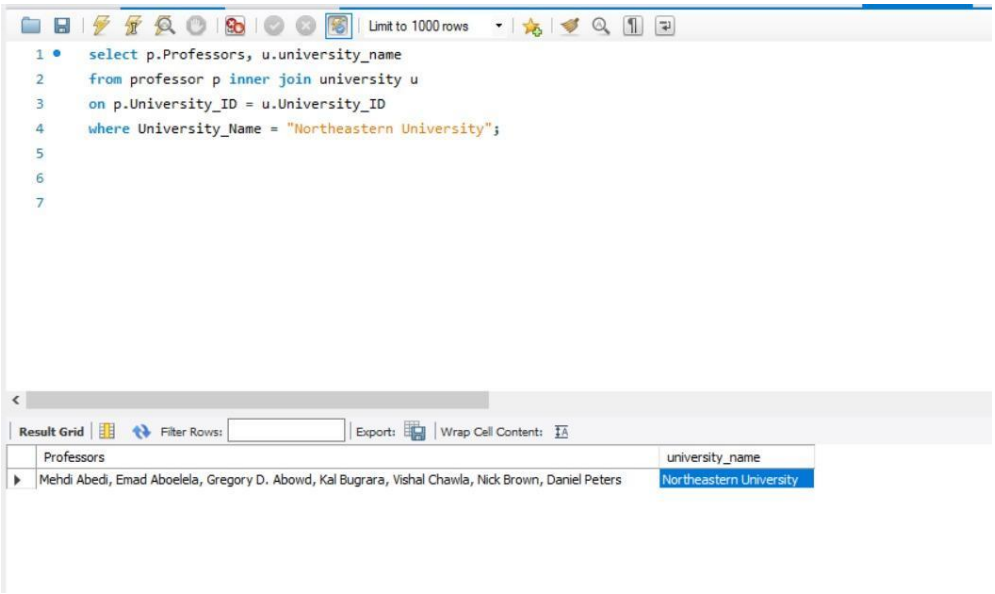


Recruiter_Name	Recruiter_Ranking	Number_of_LCA	Average_Salary	Max_Hiring_from_University
Cognizant Technology Solutions	1	12681	$92,766	Texas A&M University, Kingsville
Amazon	2	11486	$126,163	University of Washington
Tata Consultancy Services	3	9822	$93,484	Indiana University, Bloomington
Google	4	9421	$156,793	Stanford University
Microsoft	5	7329	$146,569	University of Washington
Facebook	6	6090	$175,750	Carnegie Mellon University
Ernst & Young	7	5797	$118,653	University of Texas, Arlington
Infosys	8	5649	$92,689	University of the Pacific, Stockton
Apple	9	4239	$168,416	University of Washington
Hcl America	10	4113	$99,754	San Jose State University
Accenture	11	3830	$140,893	New York University

3. **Display the list of all the professors that teach under a particular university.**
   Description: The student views the professors from a particular university
   Actor: Student
   Precondition: The student must select a university
   Steps:
   Actor Action: The student views the professor from a particular university
   System Response: The system generates a list of professors from a university
   Postcondition: The system generates all the professors

   SQL Query:

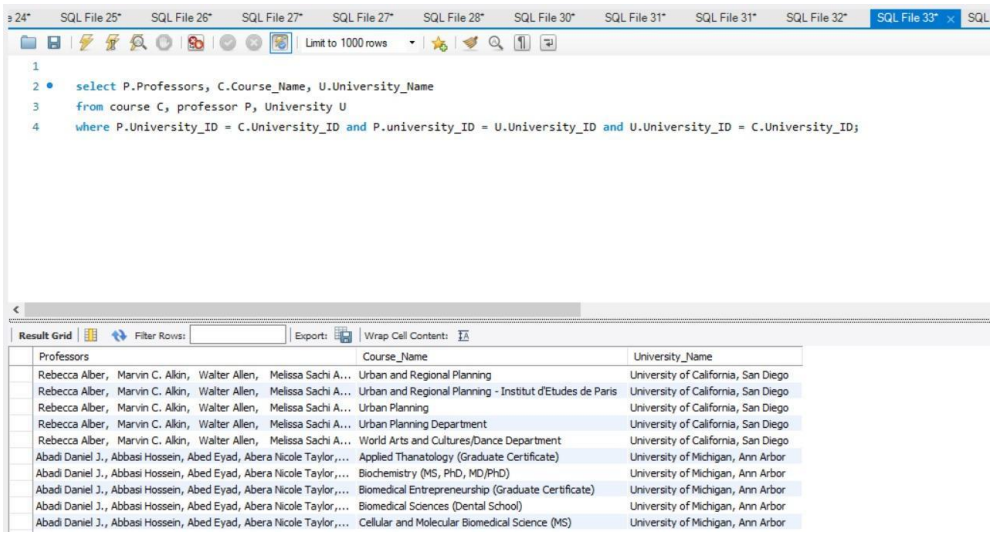
   select p.Professors, u.university_name from professor p inner join university u on p.University_ID = u.University_ID where University_Name = "Northeastern University"



4. **Display the list of all the courses offered by an university along with the list of all the professors present in the university.**
   Description: This query returns a list of professors along with their university ID and the courses offered by that university
   Actor: Student
   Precondition: The student must select a professor
   Steps:
   Actor Action: The student views the professor teaching a particular course from a university
   System Response: The system displays a list of all the professors from a university along with the professors
   Postcondition: A list is generated of all the professors at a university along with their courses

   SQL Query:

```
select P.Professors, C.Course_Name, U.University_Name
from course C, professor P, University U
where P.University_ID = C.University_ID and P.university_ID = U.University_ID
and U.University_ID = C.University_ID;
```

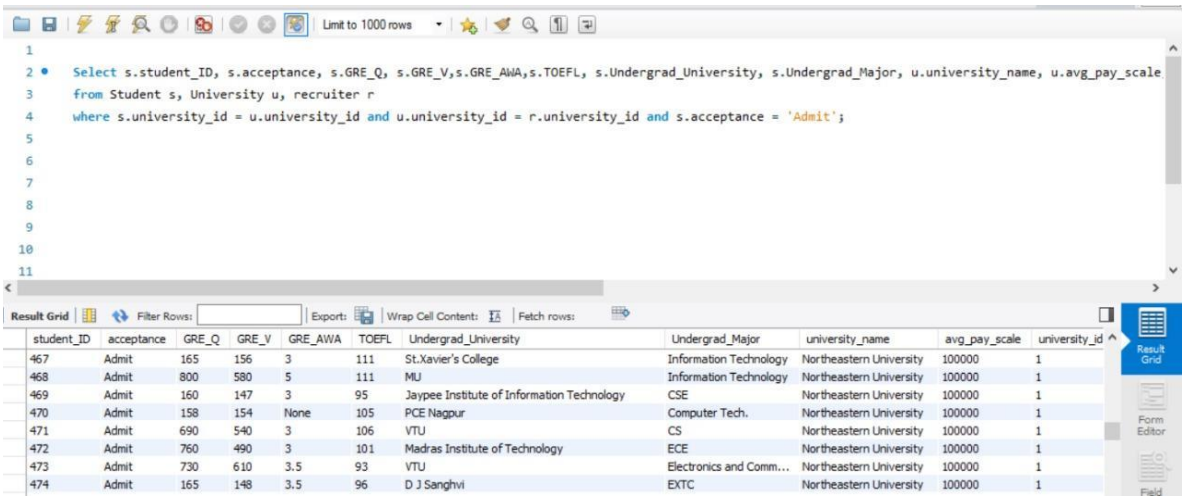


5. **Display the conditions on which a student is admitted to a university along with their graduate major, and average pay scale after graduation**

Description: Displays the entire academics of a student based on admit
Actor: Student
Precondition: The student must select a student ID
Steps:
Actor Action: The student selects a student ID and views the entire academic
System Response: The system generates a list of all the admitted students
Postcondition: A list is generated with all the amidst

SQL Query:

```
Select s.student_ID, s.acceptance, s.GRE_Q, s.GRE_V,s.GRE_AWA,s.TOEFL,
s.Undergrad_University, s.Undergrad_Major, u.university_name,
u.avg_pay_scale, r.university_id
from Student s, University u, recruiter r where s.university_id = u.university_id
and u.university_id = r.university_id and s.acceptance = 'Admit';
```



```
Select s.student_ID, s.acceptance, s.GRE_Q, s.GRE_V,s.GRE_AWA,s.TOEFL, s.Undergrad_University, s.Undergrad_Major, u.university_name, u.avg_pay_scale
from Student s, University u, recruiter r
where s.university_id = u.university_id and u.university_id = r.university_id and s.acceptance = 'Admit';
```

Result Grid

student_ID	acceptance	GRE_Q	GRE_V	GRE_AWA	TOEFL	Undergrad_University	Undergrad_Major	university_name	avg_pay_scale	university_id
467	Admit	165	156	3	111	St.Xavier's College	Information Technology	Northeastern University	100000	1
468	Admit	800	580	5	111	MU	Information Technology	Northeastern University	100000	1
469	Admit	160	147	3	95	Jaypee Institute of Information Technology	CSE	Northeastern University	100000	1
470	Admit	158	154	None	105	PCE Nagpur	Computer Tech.	Northeastern University	100000	1
471	Admit	690	540	3	106	VTU	CS	Northeastern University	100000	1
472	Admit	760	490	3	101	Madras Institute of Technology	ECE	Northeastern University	100000	1
473	Admit	730	610	3.5	93	VTU	Electronics and Comm...	Northeastern University	100000	1
474	Admit	165	148	3.5	96	D J Sanghvi	EXTC	Northeastern University	100000	1

6. **Display the list of a specific course offered by all the universities.**

Description: This query returns a value of all the universities that provide a particular course

Actor: Student

Precondition: The student must select a course

Steps:

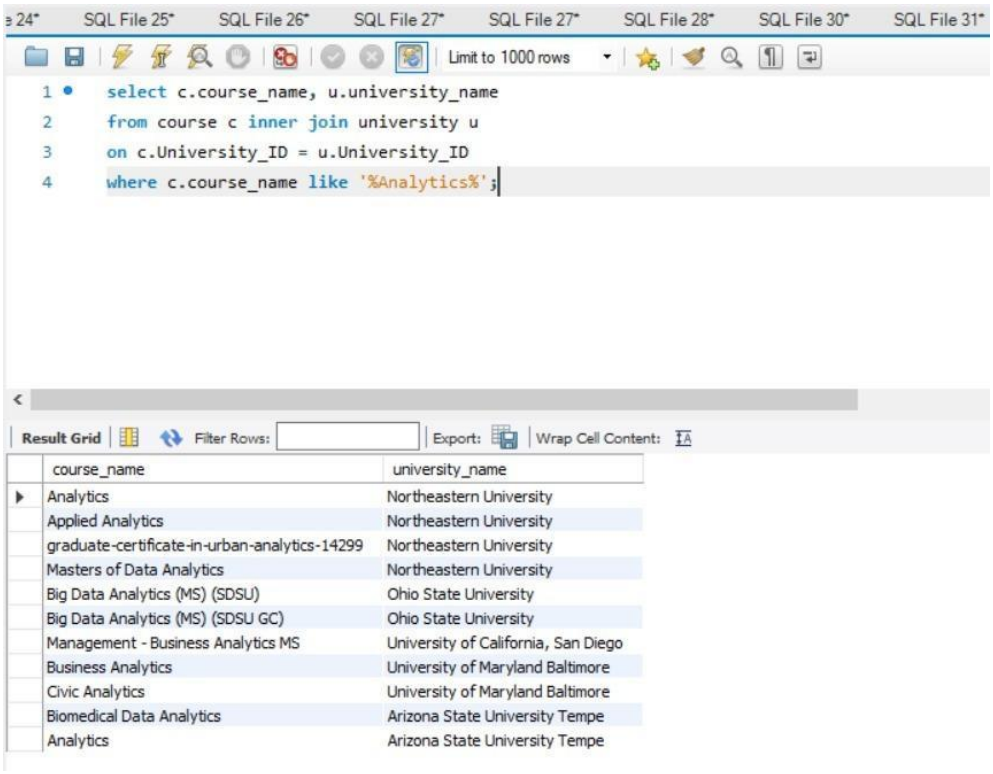
Actor Action: The student selects a particular course that he wishes to be enrolled for

System Response: The system displays a list of all the universities that provide that particular course

Postcondition: A list of all the universities that offer analytics courses is displayed

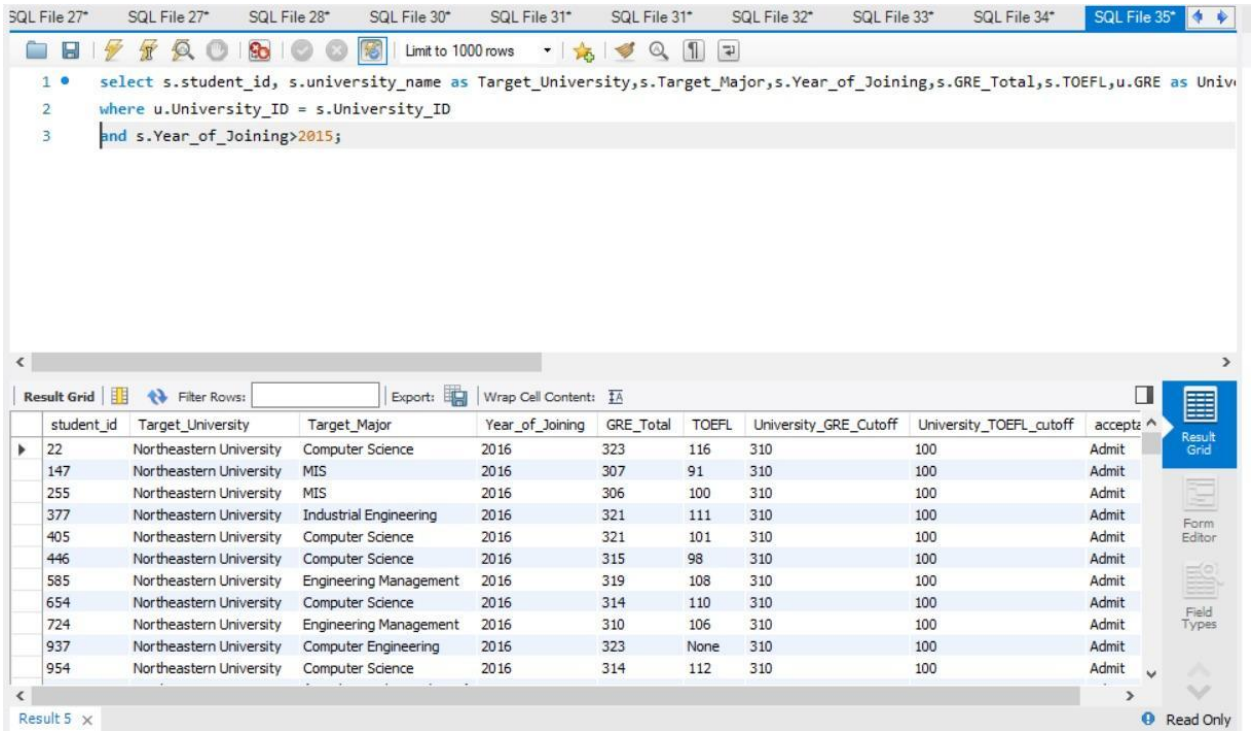
SQL Query:

```
select c.course_name, u.university_name from course c inner join university u on c.University_ID = u.University_ID where c.course_name like '%Analytics%';
```



7. **Display the data about previous admits where the year of joining >= 2015**
   Description: To view a list of all the previous admits to a university after 2015
   Actor: The student
   Precondition: The student must select a year
   Steps:
   Actor Action: The student selects the year and displays the admits
   System Response: The system displays a list of all the previous admits
   Postcondition: The list is displayed with all the previous admits

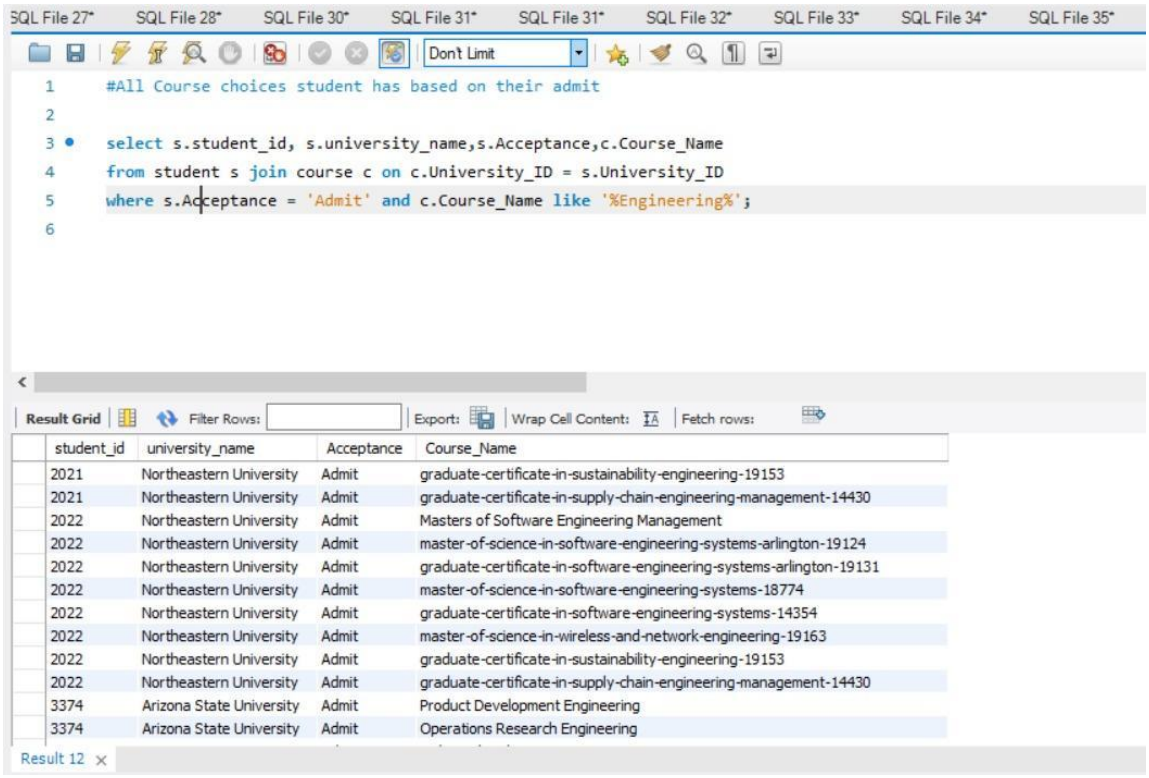
   SQL Query:
   select s.student_id, s.university_name as Target_University,s.Target_Major,s.Year_of_Joining,s.GRE_Total,s.TOEFL,u.GRE as University_GRE_Cutoff,U.TOEFL as University_TOEFL_cutoff,s.acceptance from student s, university u where u.University_ID = s.University_ID and s.Year_of_Joining>2015;



student_id	Target_University	Target_Major	Year_of_Joining	GRE_Total	TOEFL	University_GRE_Cutoff	University_TOEFL_cutoff	accepta
22	Northeastern University	Computer Science	2016	323	116	310	100	Admit
147	Northeastern University	MIS	2016	307	91	310	100	Admit
255	Northeastern University	MIS	2016	306	100	310	100	Admit
377	Northeastern University	Industrial Engineering	2016	321	111	310	100	Admit
405	Northeastern University	Computer Science	2016	321	101	310	100	Admit
446	Northeastern University	Computer Science	2016	315	98	310	100	Admit
585	Northeastern University	Engineering Management	2016	319	108	310	100	Admit
654	Northeastern University	Computer Science	2016	314	110	310	100	Admit
724	Northeastern University	Engineering Management	2016	310	106	310	100	Admit
937	Northeastern University	Computer Engineering	2016	323	None	310	100	Admit
954	Northeastern University	Computer Science	2016	314	112	310	100	Admit

8. **Display the list of courses choices based on specifications a student can select from post getting the admit.**
   Description: To view a list of all the possible choices student has
   Actor: Student
   Precondition: The student must select a course name for acceptance
   Steps:
   Actor Action: The student must select a course name
   System Response: A list is displayed with all the possible admits a student has from a university
   Postcondition: A list is displayed with all possible universities

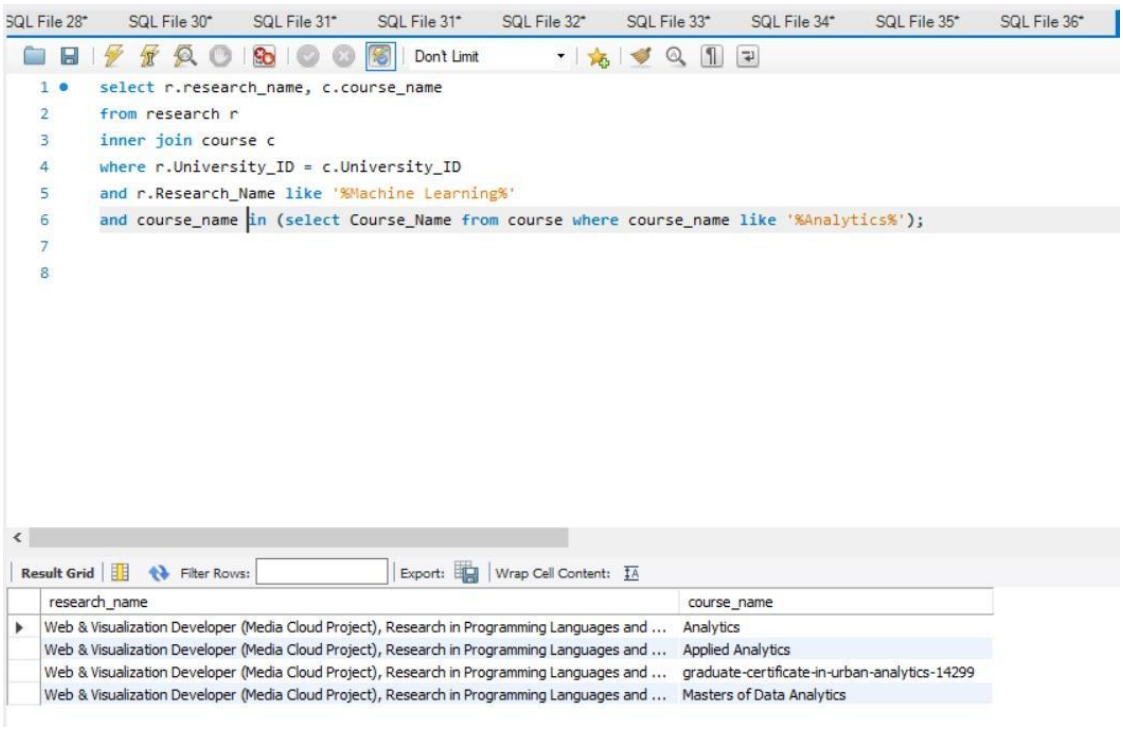
   SQL Query:
   select s.student_id, s.university_name,s.Acceptance,c.Course_Name from student s join course c on c.University_ID = s.University_ID where s.Acceptance = 'Admit' and c.Course_Name like '%Engineering%';



student_id	university_name	Acceptance	Course_Name
2021	Northeastern University	Admit	graduate-certificate-in-sustainability-engineering-19153
2021	Northeastern University	Admit	graduate-certificate-in-supply-chain-engineering-management-14430
2022	Northeastern University	Admit	Masters of Software Engineering Management
2022	Northeastern University	Admit	master-of-science-in-software-engineering-systems-arlington-19124
2022	Northeastern University	Admit	graduate-certificate-in-software-engineering-systems-arlington-19131
2022	Northeastern University	Admit	master-of-science-in-software-engineering-systems-18774
2022	Northeastern University	Admit	graduate-certificate-in-software-engineering-systems-14354
2022	Northeastern University	Admit	master-of-science-in-wireless-and-network-engineering-19163
2022	Northeastern University	Admit	graduate-certificate-in-sustainability-engineering-19153
2022	Northeastern University	Admit	graduate-certificate-in-supply-chain-engineering-management-14430
3374	Arizona State University	Admit	Product Development Engineering
3374	Arizona State University	Admit	Operations Research Engineering

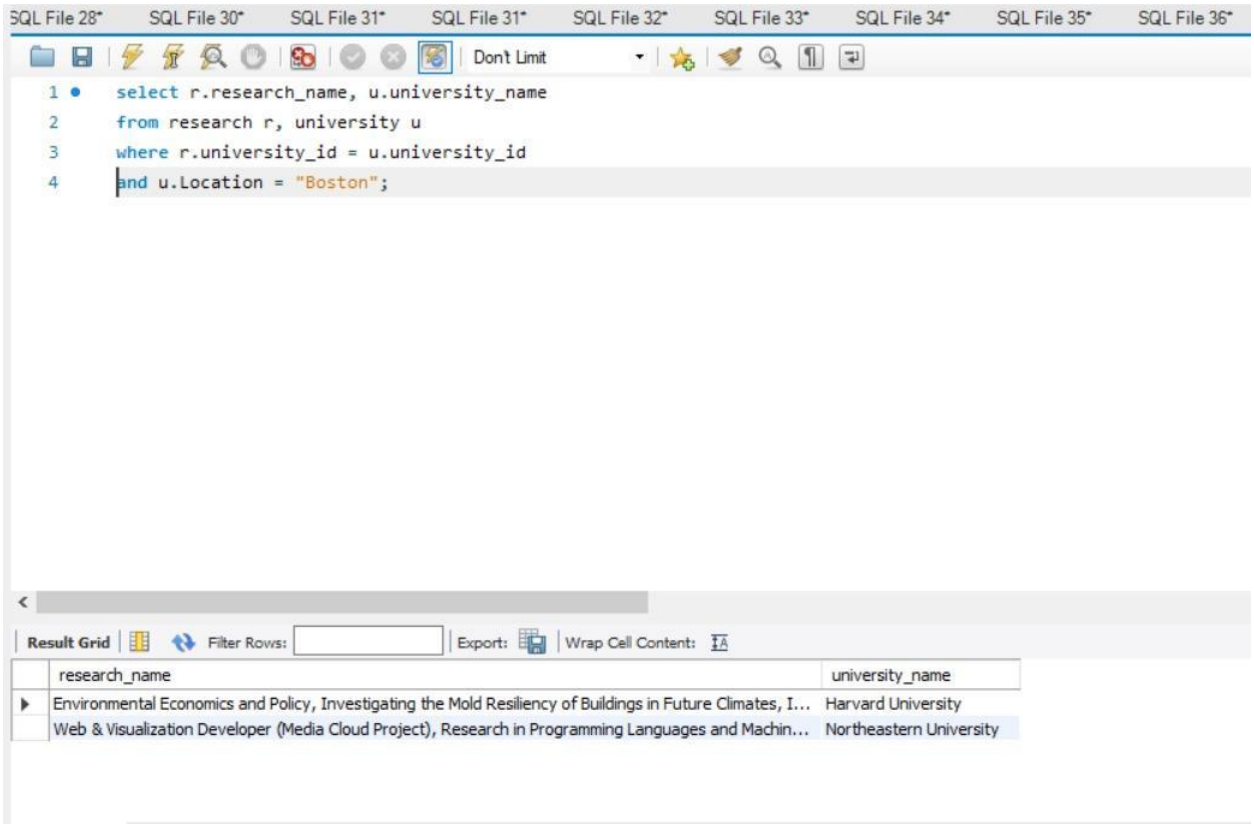
9. **Display the list of desired research opportunities available in a university based on the course selection of a student.**
   Description: To view the list of all the research in a particular course
   Actor: Student
   Precondition: The student must select a research domain say Machine Learning
   Steps:
   Actor Action: The student must select a domain for research
   System Response: The system displays a list of all the research in a particular course
   Postcondition: A list is displayed with all the research going on in a course
   SQL Query:

   select r.research_name, c.course_name from research r inner join course c where r.University_ID = c.University_ID and r.Research_Name like '%Machine Learning%' and course_name in (select Course_Name from course where course_name like '%Analytics%');



10. **Display the list of research opportunities available in an university based on the location of the university.**
    Description: To display a list of research going in a particular location in which the university exists
    Actor: Student
    Precondition: The student must select a location
    Steps:
    Actor Action: The student selects a location to look for research
    System Response: The system displays a list of universities in that location where research is going on
    Postcondition: A list is displayed with the research going in a particular location in which the university exists
    SQL Query:

    select r.research_name, u.university_name from research r, university u where r.university_id = u.university_id and u.Location = "Boston";


SQL File 28*
SQL File 30*
SQL File 31*
SQL File 31*
SQL File 32*
SQL File 33*
SQL File 34*
SQL File 35*
SQL File 36*
Don't Limit
1 select r.research_name, u.university_name
2 from research r, university u
3 where r.university_id = u.university_id
4 and u.Location = "Boston";
Result Grid
Filter Rows:
Export:
Wrap Cell Content:
research_name
university_name
Environmental Economics and Policy, Investigating the Mold Resiliency of Buildings in Future Climates, I...
Harvard University
Web & Visualization Developer (Media Cloud Project), Research in Programming Languages and Machin...
Northeastern University