CUNY HUB

CSCI 435: Jonnathan Brito, Nursima Donuk, Alfredo Morales

IBM Cloud sponsored by Power Systems
Academic Initiative

About Our Project

- The major issue we wanted to tackle was **lack of information.**

- Attending NYC high schools we noticed there isn't somewhere we can find information on CUNY schools all in one place. *It's hard to find what you're looking for if you don't know where to look for it.*

- **Initial Objective:** Upon finding an appropriate dataset, we had planned to construct a hub for CUNY information.

- However, our data, while quite extensive, lacked the necessary components to assemble it how we imagined it in its entirety.

- **Revised Objective:** Generate data and combine it with what was available to us, perform normalizations, run some queries which reinforce class objectives and present our vision.

The Data

- Our data was obtained through various CUNY websites, NY Payroll system searches, and some was generated using R.

Our Attributes

CONTACT.CSV -----
- SchoolId
- Campus
- CampusWebsite
- PhoneNumber
- Address
- Latitude
- Longitude

EXPENSES.CSV ----
- SchoolId
- Resident
- Transp
- BookSpl
- Lunch
- RoomBrd
- PerCr

COURSES.CSV -----
- DeptId
- CourseNo
- CourseCr

DEPARTMENTS.CSV -
- DeptId
- SchoolId
- DeptName

FACULTY.CSV -----
- FactultyId
- Name
- Rating
- Email
- StartYear

DEGREES.CSV -----
- SchoolId
- ProgramName
- AwardName
- DegreeDept

DEPSTAUGHT.CSV ---
- FacultyId
- DeptId
- Year

HONORS.CSV -------
- SchoolId
- HonorsProgram
- Link

EXTRA.CSV -------------
- SchoolId
- AvgFinAid
- Type
- Size
- TotalUndergraduate
- TotalGraduates
- Gender%
- AvrgAgeAll
- AvrgAgeFull
- PartTime%
- Ethnicity%
- InState%
- OutOfState%
- StudentFacultyRate
- MedianSAT
- AdmissionType
- AdmittedApplicants%

Pre-Normalization Table Definition

School_Name	Hunter College	varchar(30)
School_address:	695 Park Ave, New York, NY 1006	varchar(30)
School_phone_number	2127724000	int
Course	CSCI 435 Data Base Management	int
Department	Computer Science	varchar(20)
Number of Students	23,018	int
Number of Professors	240	int
Room and Board	$4,750	float
Number of Majors	83	int
Cost_per_credit	462	int
Professor	Raman Kannan	varchar(30)
Majors	Bachelor of Arts in Computer Scier	int
Minors	Minor in Computer Sceince	int
State Resident Expenses	$9,921	float
Rate My Professor	4.2	float

Normalization => 1NF

First Normal Form Rules: Atomic columns, unique records

- *Address* and *professor_name* columns and were not atomic initially, therefore we split in a way that complies with 1NF.

Criteria:
- All Attributes must be Atomic
- Must eliminate Vertical and Horizontal Repetition
- Must Identify a Pk

School_Name	Hunter College	Baruch College	Yes	N/A
School_ID	1	2	Yes	N/A
School_address:	695 Park Ave, New York, NY 1006	55 Lexington Ave, New York, NY 10010, Un	No	street
				city
				state
				zip
				country
School_phone_number	2127724000	6463121000	Yes	N/A
Course	CSCI 435 Data Base Management	BUSI 101 Introduction to Accounting	No	Course ID (CID)
				Course Name
Department:	Computer Science	Accounting	No	Majors
				Minors
				Programs
				Professors
Number of Students	23,018	18,090	Yes	N/A
Number of Professors	240	192	Yes	N/A
Number of Majors	83	35	Yes	N/A
Cost_per_credit	462	436	Yes	N/A
Professor	Raman Kannan	Coleson Robert	No	Professor Name
				Professor ID(IPD)
				Professor Rating
				Research Projects
				Date they started teaching at CUNY
				Courses taught
Majors	Bachelor of Arts in Computer Scien	Accountancy of BBA	No	Courses Offered
				Credit Requirement
				Number of Electives Required
				Required Courses

Normalization => 2NF

Second Normal Form Rules: Be in 1NF, Single Column Primary Key.

- SchoolID is the **primary key** of the Contact table.
- SchoolID is also the **foreign key** of the Degrees, Departments, Expenses, Extra, and Honors tables.

- DeptId is the **primary key** of the Departments table.
- DeptId is the **foreign key** of the Deptstaught table.

- FactultyId is the **primary key** of the Faculty table.
- FactultyId is the **foreign key** of the Deptstaught table.

- **No PARTIAL DEPENDENCY**

Normalization => 3NF

- **Third Normal Form Rules:** Dataset is in 2NF and there is no transitive dependency.
  - An attribute can't depend on another attribute

	A	B	C	D	E	F	G	H	I	J
1										
2	CONTACT.CSV									
3	SchoolId	Campus	CampusWebsite	PhoneNumber	Street	City	State	Zip	Latitude	Longitude
4	1	Baruch College	http://baruch.cuny.edu	(646) 312-1000	55 Lexington Ave at 24th Street	New York	NY	10010-2313	40.740977	-73.984252
5	2	BMCC	http://bmcc.cuny.edu	(212) 220-1265	199 Chambers Street	New York	NY	10007-1044	40.717367	-74.012178
6	3	Bronx Community College	http://bcc.cuny.edu	(718) 289-5100	2155 University Avenue	Bronx	NY	10453	40.856673	-73.910127
7	4	Brooklyn College	http://brooklyn.edu	(718) 951-5001	2900 Bedford Avenue	Brooklyn	NY	11210-2850	40.630276	-73.955545
8										
9	COURSES.CSV				DEGREES.CSV					
10	DeptId	CourseNo	CourseCr		SchoolId	ProgramName		AwardName	CIP201TITLESHORT	
11	1	640	4		1	Accountancy		M.B.A.	Accounting.	
12	1	400	3		1	Accounting		B.B.A.	Accounting.	
13	1	330	3		1	Statistics and Quantitative Modeling		B.B.A.	Business Statistics.	
14	1	620	3		1	Statistics		M.B.A.	Business Statistics.	
15										
16	DEPARTMENTS.CSV				DEPSTAUGHT.CSV					
17	DeptId	SchoolId	DeptName		FacultyId	DeptId	Year			
18	1	1	ITAL		1	181	2019			
19	2	1	LATED		2	666	2015			
20	3	1	NURS		2	620	2007			
21	4	1	FILM		2	176	2020			
22										
23	EXPENSES.CSV									
24	SchoolId	Resident	Transp	BookSpl	Lunch	Prsnl	RoomBrd	PerCr		
25	1	10294	1615	1380	982	1505	3211	499		
26	2	10098	1643	578	931	1431	4024	436		
27	3	9661	1724	745	1111	1692	4874	416		
28	4	10693	1616	807	1131	1471	3626	468		
29										

Normalization => 3NF Continued...

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
29														
30	EXTRA.CSV													
31	SchoolId	AvgFinAid	Type	Size	TotalUndergraduates	TotalGraduates	Women%	Men%	AvrgAgeAll	AvrgAgeFull	PartTime%	Hispanic/Latino%		
32	1	10615	Senior Colleges	Large	15482	3197	0.47	0.53	23	22	0.2	0.26		
33	2	5850	Community Colleges	Large	25500		0.57	0.43	24	22	0.26	0.44		
34	3	5977	Community Colleges	Medium	10040		0.56	0.44	28	28	0.38	0.62		
35	4	10322	Senior Colleges	Medium	14970	2841	0.57	0.43	24	22	0.21	0.25		
36	Asian%	White%	Black/AfricanAmerica	TwoOrMore%	NonResidentAlien%	InState%	OutOfState%	StudentFacultyR	MedianSAT	AdmissionType	AdmittedApplicants%			
37	0.34	0.2	0.09	0.02	0.1	0.97	0.03	19	1230	very selective	0.43			
38	0.11	0.09	0.26	0.02	0.07	0.97	0.03	22	890	open admission	0.93			
39	0.03	0.02	0.28	0.01		0.99	0.01	19		open admission				
40	0.21	0.28	0.2	0.02	0.03	0.98	0.02	16	1130	very selective	0.45			
41														
42	FACULTY.CSV							HONORS.CSV						
43	FacultyId	LastName	FirstName	Rating	Email	StartYear		SchoolId	HonorsProgram			Link		
44	1	Gardner	Kevin	3.8	kegardner@email.cuny.edu	2013		1	The Baruch Honors Program			https://blogs.baruch.cuny.edu/honors/		
45	2	Hotzler	Russell K	5	ruhotzler@email.cuny.edu	2002		2	The Honors Program at BMCC			https://www.bmcc.cuny.edu/academics/		
46	3	Schaffler	Mitchell	3.5	mischaffler@email.cuny.edu	2014		3	The Honors Program at Bronx Community College			https://bcc-cuny.digication.com/honors/Home		
47	4	Vorosmarty	Charles	3.4	chvorosmarty@email.cuny.edu	2006		4	The Coordinated B.A.-M.D. Program			http://www.brooklyn.cuny.edu/web/academics/h		
48														
49														

Criteria:

NO
TRANSITIVE
DEPENDENCY

Non-Transitive Examples:
Person ID- First Name, Last Name

A person's last name is based on their father's last name and

Query Demonstrations

Number of Courses per School

```
mysql> SELECT contact.campus campus, B.cnt numcrs FROM contact JOIN (SELECT schid, count(*) cnt FROM (SELECT departments.schid schid,
 courses.courseno crs FROM courses JOIN departments ON departments.deptid = courses.deptid) A GROUP BY schid) B ON B.schid = contact.
schid;
+-------------------------------------------+--------+
| campus                                    | numcrs |
+-------------------------------------------+--------+
| Baruch College                           | 1320 |
| Borough of Manhattan Community College   | 1325 |
| Bronx Community College                  | 1294 |
| Brooklyn College                         | 1492 |
| College of Staten Island                 | 1309 |
| Guttman Community College                | 1340 |
| Hostos Community College                 | 1262 |
| Hunter College                           | 1423 |
| John Jay College of Criminal Justice     | 1364 |
| Kingsborough Community College           | 1237 |
| LaGuardia Community College              | 1460 |
| Lehman College                           | 1364 |
| Medgar Evers College                     | 1322 |
| New York City College of Technology      | 1451 |
| Queensborough Community College          | 1479 |
| Queens College                           | 1462 |
| The City College of New York             | 1477 |
| York College                             | 1338 |
+-------------------------------------------+--------+
18 rows in set (0.05 sec)
```

Query Demonstrations Continued

Average Faculty Rating for each School

```
mysql> SELECT camp, AVG(rating) avg FROM (SELECT B.camp camp, faculty.rating rating FROM faculty JOIN (SELECT contact.campus camp, A.factid
factid FROM contact JOIN (SELECT departments.schid schid, depsTaught.facultyid factid FROM departments JOIN depsTaught ON departments.deptid
= depsTaught.deptid) A ON A.schid = contact.schid) B ON B.factid = faculty.facultyid) C GROUP BY camp ORDER BY avg DESC;
+-----------------------------------------+--------------------+
| camp                                    | avg                |
+-----------------------------------------+--------------------+
| Medgar Evers College                    | 3.141254523522317  |
| Guttman Community College               | 3.1329100529100486 |
| LaGuardia Community College             | 3.1312722948870424 |
| New York City College of Technology     | 3.0977491961414807 |
| Brooklyn College                        | 3.094543546694648  |
| Hunter College                          | 3.074765380604797  |
| Queensborough Community College         | 3.072549019607843  |
| Baruch College                          | 3.0548317046688345 |
| York College                            | 3.053070683661639  |
| Hostos Community College                | 3.048753117206982  |
| Kingsborough Community College          | 3.0455405405405385 |
| College of Staten Island                | 3.045328719723187  |
| Bronx Community College                 | 3.0404255319148965 |
| Lehman College                          | 3.0236872812135354 |
| The City College of New York            | 3.014591836734694  |
| Queens College                          | 3.004042553191488  |
| Borough of Manhattan Community College  | 2.979975124378109  |
| John Jay College of Criminal Justice    | 2.95809426229508   |
+-----------------------------------------+--------------------+
18 rows in set (0.06 sec)
```

Query Demonstrations Continued

Number of Degrees Available at each School

```
mysql> SELECT contact.campus sch, A.cnt deg FROM contact JOIN (SELECT schid, count(*) cnt FROM degrees GROUP BY schid) A ON A.schid = contact.schid;
+-----------------------------------------+-----+
| sch                                     | deg |
+-----------------------------------------+-----+
| Baruch College                          |  85 |
| Borough of Manhattan Community College  |  55 |
| Bronx Community College                 |  50 |
| Brooklyn College                        | 276 |
| College of Staten Island                |  96 |
| Guttman Community College               |   5 |
| Hostos Community College                |  33 |
| Hunter College                          | 316 |
| John Jay College of Criminal Justice    |  97 |
| Kingsborough Community College          |  60 |
| LaGuardia Community College             |  56 |
| Lehman College                          | 212 |
| Medgar Evers College                    |  33 |
| New York City College of Technology     |  76 |
| Queensborough Community College         |  54 |
| Queens College                          |  75 |
| The City College of New York            | 247 |
| York College                            |  67 |
+-----------------------------------------+-----+
18 rows in set (0.00 sec)
```

Query Demonstrations Continued

Most Offered Course Number

```
[mysql> SELECT courseno, count(*) cnt FROM courses GROUP BY courseno ORDER BY cnt DESC LIMIT 1;
+----------+-----+
| courseno | cnt |
+----------+-----+
|      470 | 381 |
+----------+-----+
1 row in set (0.03 sec)
```

Average Number of Departments Across All School Districts

```
[mysql> SELECT AVG(cnt) FROM (SELECT schid, count(*) cnt FROM (SELECT schid, deptname FROM (SELECT depart]
ments.schid schid, departments.deptname deptname, courses.courseno courseno FROM departments JOIN course
s ON departments.deptid = courses.deptid) A GROUP BY schid, deptname) B GROUP BY schid) C;
+----------+
| AVG(cnt) |
+----------+
|  46.5556 |
+----------+
1 row in set (0.04 sec)
```

Query Demonstrations Continued

Total Average Expenses for Each School

```
[mysql> SELECT contact.campus, A.total FROM contact JOIN (SELECT schid, resident+transp+bookspl+lunch+prsnl+rmbrd+credt total FROM exp]
enses) A ON A.schid = contact.schid ORDER BY total;
+------------------------------------------+-------+
| campus                                   | total |
+------------------------------------------+-------+
| LaGuardia Community College              | 17422 |
| The City College of New York             | 17604 |
| College of Staten Island                 | 18252 |
| Queensborough Community College          | 18766 |
| Borough of Manhattan Community College   | 19141 |
| Medgar Evers College                     | 19165 |
| Baruch College                           | 19486 |
| Lehman College                           | 19497 |
| York College                             | 19622 |
| John Jay College of Criminal Justice     | 19681 |
| Brooklyn College                         | 19812 |
| Guttman Community College                | 19932 |
| Queens College                           | 20047 |
| New York City College of Technology      | 20055 |
| Bronx Community College                  | 20223 |
| Hostos Community College                 | 20443 |
| Hunter College                           | 20503 |
| Kingsborough Community College           | 22034 |
+------------------------------------------+-------+
18 rows in set (0.00 sec)
```

Acknowledgments

- We would like to thank Dr. Raman Kannan for guiding us throughout the trajectory of the project.

- We would also like to thank IBM Cloud Power Systems for supporting the Academic Initiative and propagating our understanding of Databases.

- We all thoroughly enjoyed this class.

Thank you for your time!

Us when
presenting :)

=>