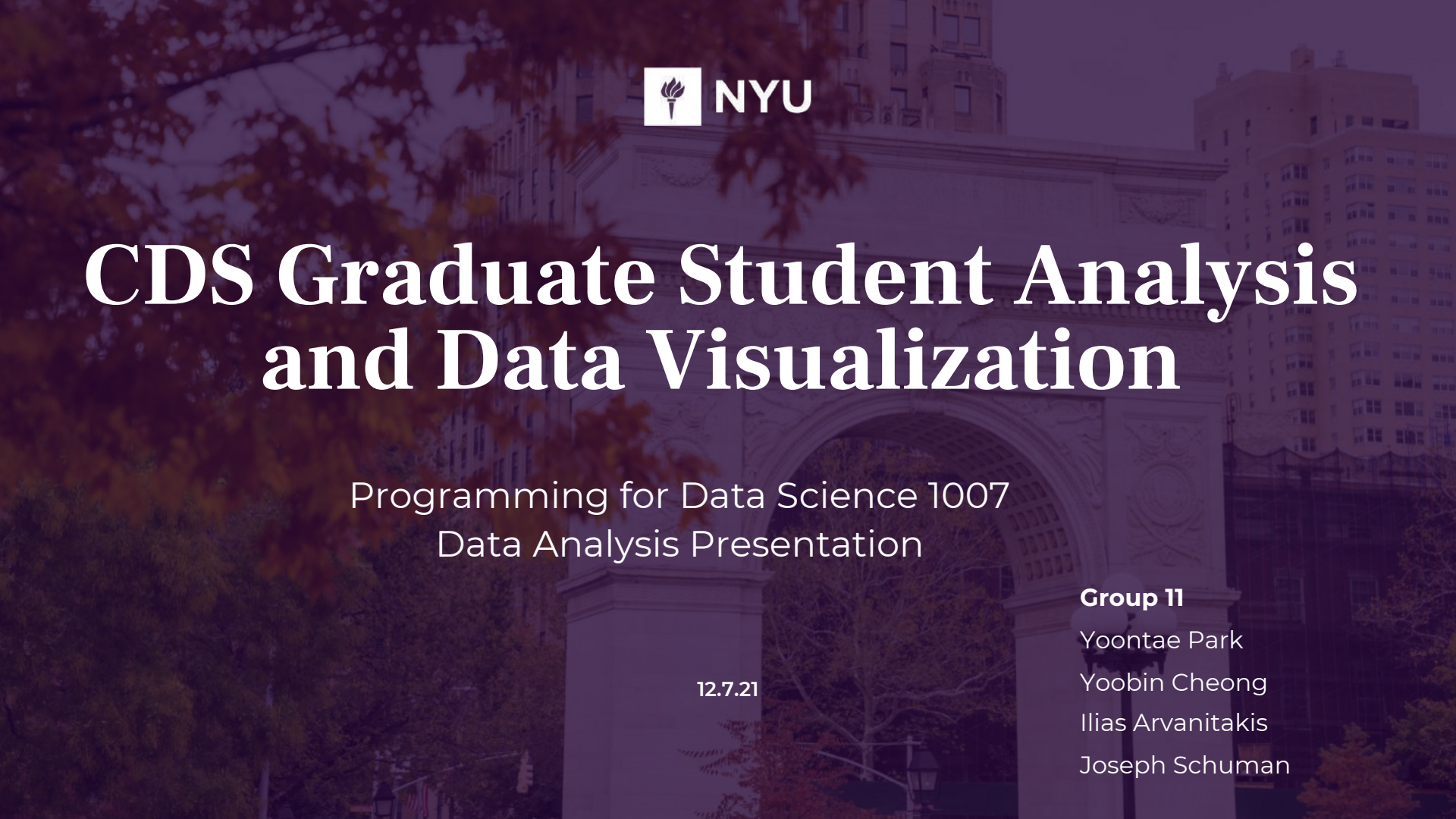


NYU

CDS Graduate Student Analysis and Data Visualization

Programming for Data Science 1007
Data Analysis Presentation

12.7.21

Group 11
Yoontae Park
Yoobin Cheong
Ilias Arvanitakis
Joseph Schuman

Objective

Develop deeper insight for CDS administration, faculty, and students

- Develop an intuition of CDS students, both current and former
- Provide analysis and data visualization report to CDS administration and faculty

Student Backgrounds: CDS Webpage

Standardized Tests

We require that students submit standardized tests scores for the GRE. There are no exceptions: we do not accept "out of date" scores; nor do we accept scores of other, similar tests; nor do we allow waivers (regardless of previous educational attainment or circumstance).

In addition to sending your official scores to the Graduate School of Arts and Science please upload a PDF of the unofficial scores, which are made available upon completion of your test, to the "Additional Information Section" of your application.

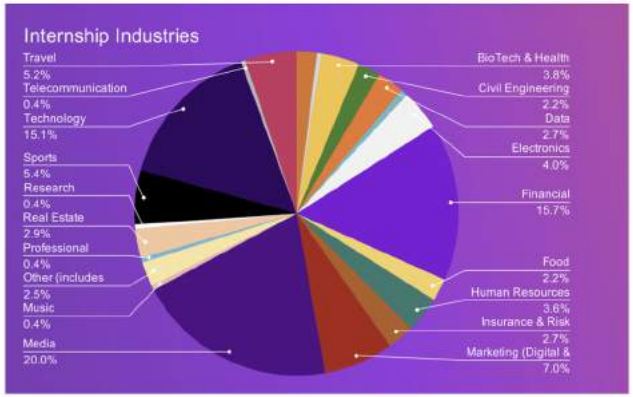
We wish to emphasize that *we have no set minimums for the GREs, and we consider the totality of an application when making a decision about admission.* Nonetheless, to the extent that it is helpful to give applicants a sense of things, what follows are the averages for the current cohort of MSDS students:

Average GRE Verbal: 159.3 (80th percentile)
Average GRE Quantitative: 167.4 (90th percentile)
Average GRE Analytical: 4.14 (61st percentile)
Average TOEFL (where required): 111

Career Outcomes: CDS Webpage

Our Students

INTERNSHIP INDUSTRY STATISTICS



Since 2016, students have completed internships in various industries.

NYU

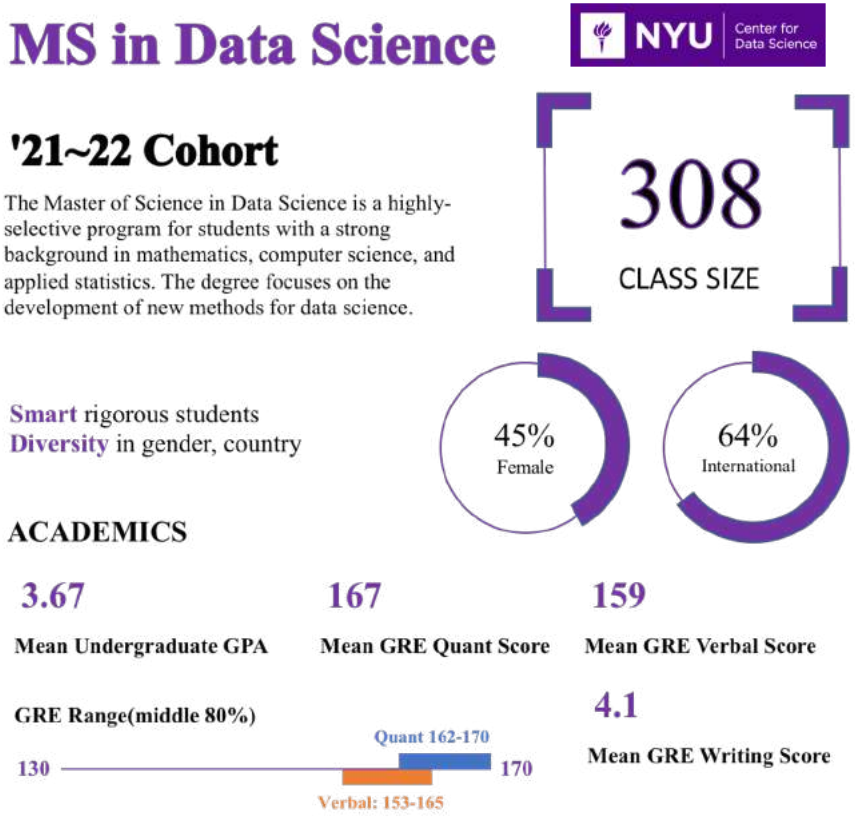
Visualizing Key Findings (Class Analytics)

GRE Mean Scores (percentile):
Quantitative: 167 (90%),
Verbal: 159 (80%),
Writing: 4.1 (62%),

Mean GPA:
3.67

Gender Diversity:
Male: 55%, Female: 45%

Geographic Diversity:
International: 64%, Domestic: 36%



Visualizing Key Findings (Career Outcomes)

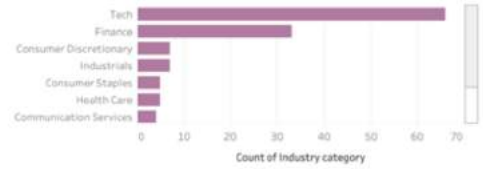
Students primarily pursue careers as Data Scientists in Tech/Finance in NY/WA/MA

Career Outcomes - Summer internship

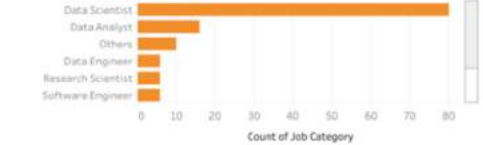
Summer(Raw)

Summer 2021	Tech	IBM	3
		Amazon	3
		Facebook	3
		Adobe	2
		Bombora	1
		eBay	2
		NVIDIA	1
		Oracle	1
		Asana	1
		FarFetch	1
		HP	1
		LinkedIn	1
		MobilityWare	1
		Software Tools Re..	1
	Finance	JP Morgan	2
		Citadel	3
		Fidelity	1

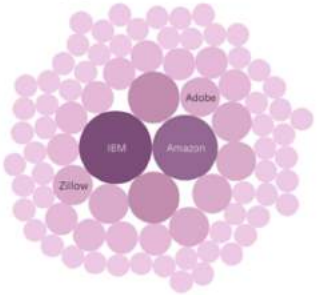
Summer(Industry)



Summer(Job_cat)



Summer(Organization)



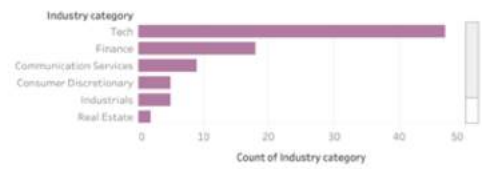
Summer(Map)

Career Outcomes - Full time offer

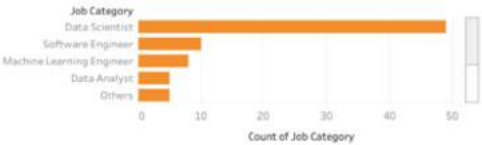
Career(Raw)

Career(20-21)	Tech	Amazon	3
		Apple	2
		Google	1
		MITRE	2
		Oracle	2
		Bombora	1
		DiDi	1
		Verisk	1
		Alpha Roc	1
		Hinge	1
		Kuai	1
	Finance	JP Morgan	1
		Prudential	2
		American Express	2
		Fidelity	1
		Ernst & Young	1
		PIMCO	1

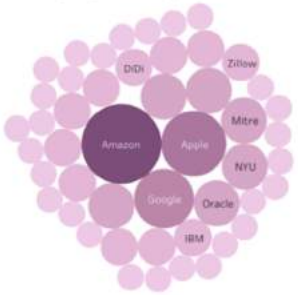
Career(Industry)



Career(Job_cat)



Career(Organization)



Career(Map)

Overview of our dataset

- **Two types of datasets**
  - Student background profile & Student Career outcomes
- **Data Characteristics**
  - Student background information:
    - Legal sex, US citizenship, GPA, GRE, Size of dataset (number of rows and column, Years of students etc.
  - Career outcomes: Company name, job title, and job location
- **Data Source information**: CDS internal data
- **NOTE:** Any data that can be used to identify an individual is NOT provided. (Name, DOB, etc)

	Legal Sex	US Cit Stat	Foreign Cit	Undergraduate GPA	TOEFL	Q	Q%	V	V%	W	W%
0	F	1	NaN	3.42	NaN	144.0	15.0	144.0	22.0	4.0	55.0
1	F	4	CHN	3.70	NaN	167.0	89.0	145.0	25.0	3.0	14.0
2	M	4	CHN	3.79	NaN	170.0	96.0	146.0	28.0	3.0	14.0
3	M	1	OTHER	3.71	NaN	160.0	72.0	147.0	32.0	4.5	80.0
4	M	4	OTHER	3.60	NaN	163.0	80.0	148.0	36.0	3.0	14.0

NYU

Data pre-processing

- **Current student profile (679 rows, 15 columns)**

  Initial data validation:

  Null values: Dropped row wise for categorical data elements (Gender, Citizenship Status)

  Converted to means for numerical data elements (GPA, GRE)

  Eliminated duplicated rows

  Additional data enhancement: conversion of different scaled data elements

  Converted from 100 point scale to 4.0 scale for GPA

  Data supplementation:

  Calculated GRE total scores

	MS Enrollment	Legal Sex	US Cit Stat	Resident Status	Foreign Cit	Undergraduate GPA	Converted GPA	Q	Q%	V	V%	W	W%	Q+V	TOEFL
0	2021-2022	F	1.0	Citizen	US	3.42	3.42	144.0	15.0	144.0	22.0	4.0	55.0	288.0	NaN
1	2021-2022	F	4.0	International	CHN	3.7	3.70	167.0	89.0	145.0	25.0	3.0	14.0	312.0	NaN
2	2021-2022	M	4.0	International	CHN	3.79	3.79	170.0	96.0	146.0	28.0	3.0	14.0	316.0	NaN
3	2021-2022	M	1.0	Citizen	OTHER	3.71	3.71	160.0	72.0	147.0	32.0	4.5	80.0	307.0	NaN
4	2021-2022	M	4.0	International	OTHER	3.6	3.60	163.0	80.0	148.0	36.0	3.0	14.0	311.0	NaN

Data pre-processing

- **Career outcomes (221 rows, 6 columns)**

  Initial data validation

  Handled Null values: Dropped row wise for categorical data elements (Company name, Job title)

  Duplicated rows were not eliminated (It makes sense many students to have the same job)

  Additional data enhancement:

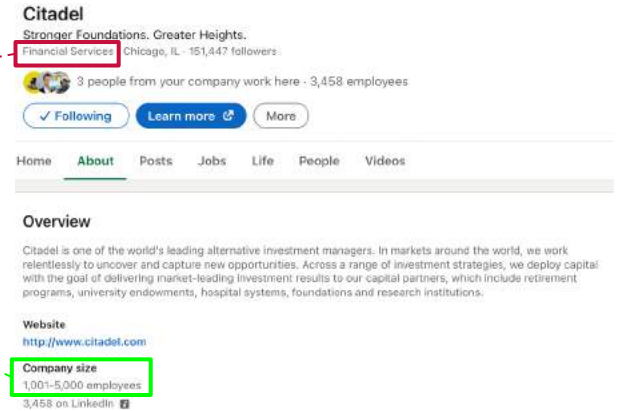
  For unidentified job location the HQ was used

  Merged typos or abbreviate names for companies into a single name

  Data supplementation:

  Created a new column from Linkedin that contained industry, company size, and HQ location

	Semester	Organization_cleaned	Industry_category	Company_size	Final_loc	Job_Category
0	Summer 2021	Citadel	Finance	1,001-5,000	IL	Data Engineer
1	Summer 2021	Citadel	Finance	1,001-5,000	IL	Data Engineer
2	Summer 2021	Citadel	Finance	1,001-5,000	IL	Data Engineer
3	Career(18-19)	Citadel	Finance	1,001-5,000	NY	Quantitative Analyst
4	Summer 2021	Adobe	Tech	10,001+	CA	Data Scientist



Results: Student Academic performance

Class Statistics (cumulative over 4 years)

- Class Statistics (Across four academic class years: 2019 - 2022) :

  Mean GRE score: Quantitative: 167, Verbal: 159, Writing: 4.1

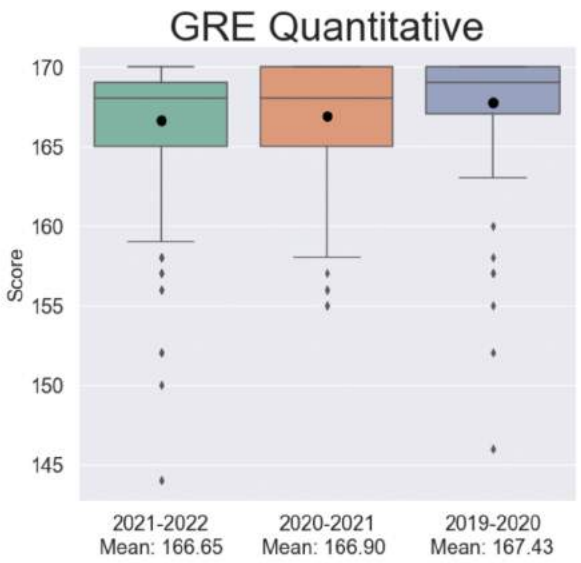
  Mean GPA 3.67 (std 0.23)

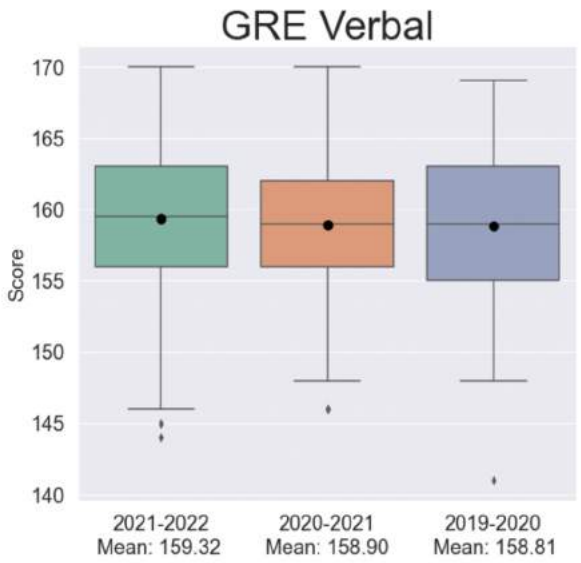
Class Statistics: GRE, GPA score

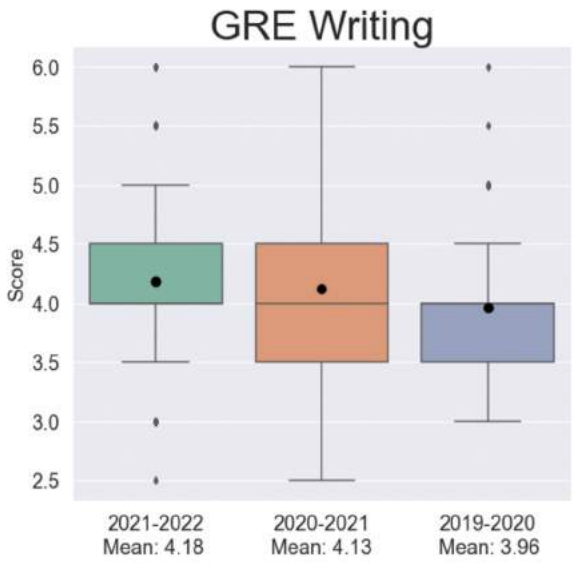
	US Cit Stat	Converted GPA	Q	Q%	V	V%	W	W%	Q+V	TOEFL
count	679.00	679.00	679.00	679.00	679.00	679.00	679.00	679.00	679.00	79.00
mean	2.96	3.67	166.95	90.04	159.08	79.94	4.12	61.70	326.03	108.59
std	1.41	0.23	3.70	9.38	4.85	14.92	0.64	22.62	6.62	3.87
min	1.00	2.54	144.00	15.00	141.00	15.00	2.50	6.00	287.00	102.00
25%	1.00	3.57	166.00	88.00	156.00	73.00	3.50	42.00	323.00	106.00
50%	4.00	3.68	168.00	94.00	159.00	83.00	4.00	59.00	327.00	108.00
75%	4.00	3.83	170.00	96.00	162.00	91.00	4.50	82.00	330.00	110.00
max	4.00	4.00	170.00	98.00	170.00	99.00	6.00	99.00	340.00	120.00

Results: Student Academic performance

Class Statistics (by academic year)





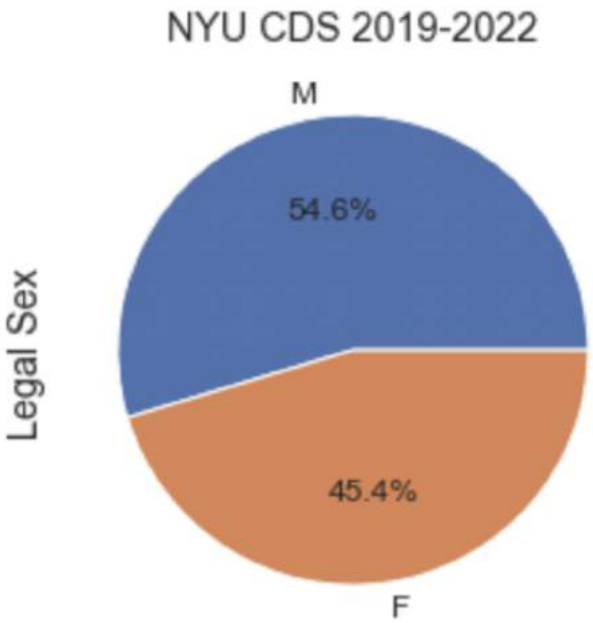


NYU

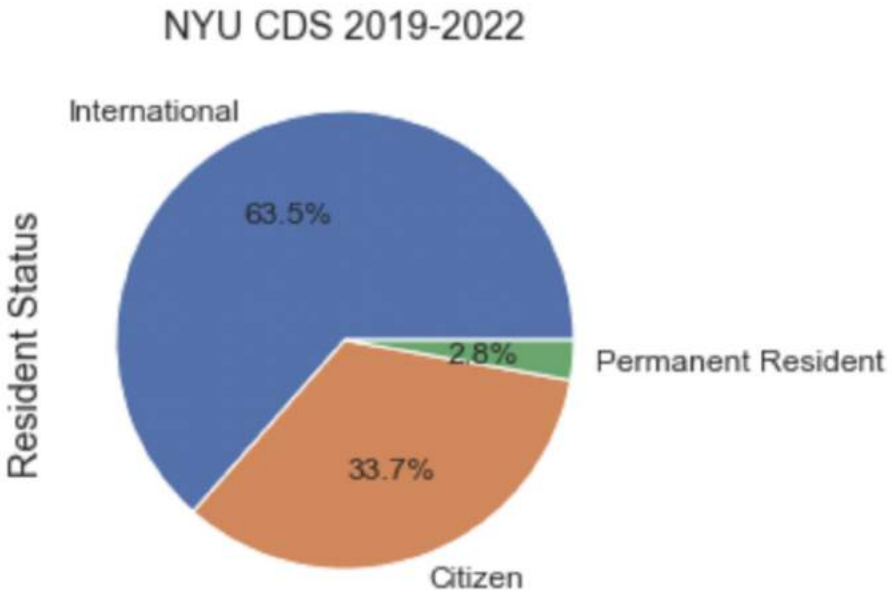
Results: CDS Student Diversity

Class Statistics (cumulative over 4 years)

Male: 54.6% , Female: 45.4%



International: 63.5%, Domestic: 36.5%



Results: Student Academic performance

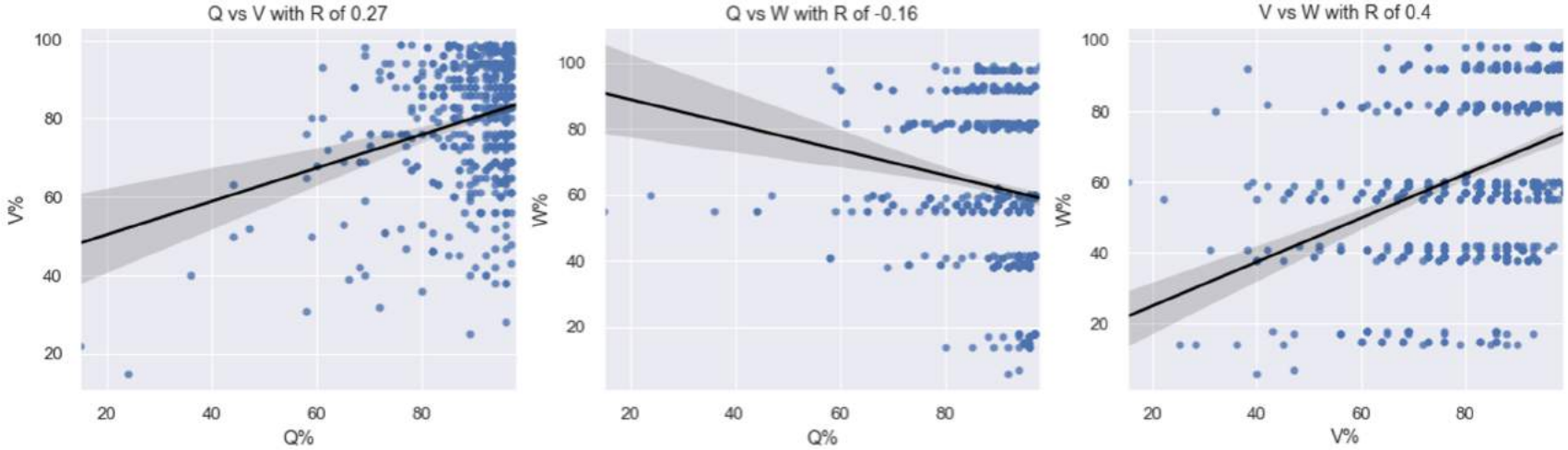
Regression Analysis

- Regression, correlation between scores:

  GRE Quant and Verbal are positively correlated with a coefficient of 0.27

  GRE Quant and Writing are negatively correlated with a coefficient of -0.16

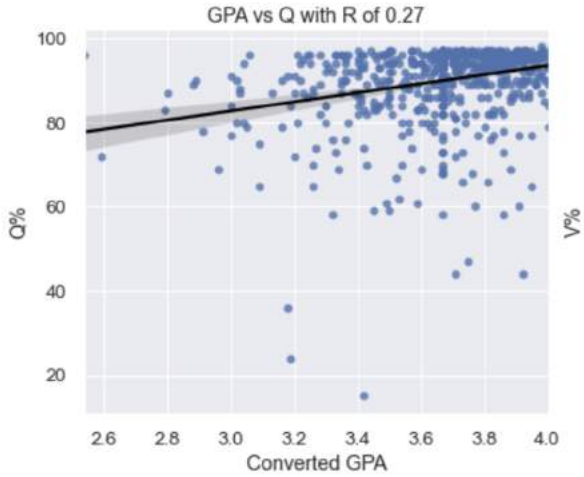
  GRE Verbal and Writing are positively correlated with a coefficient of 0.40

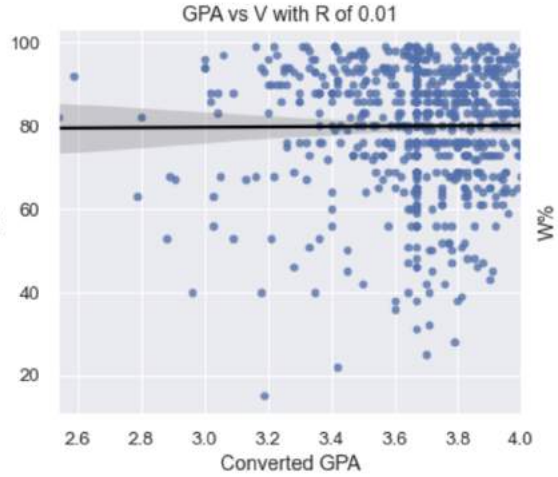


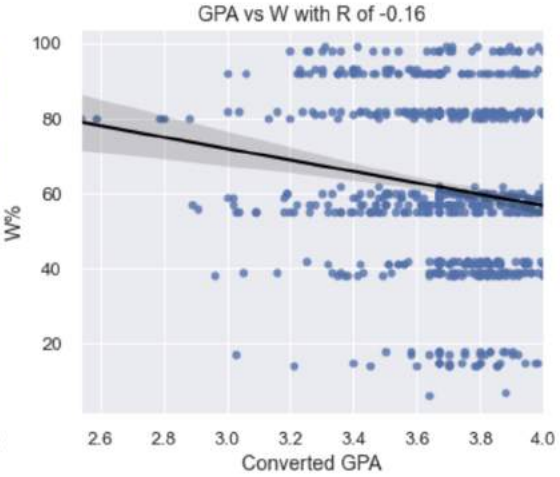
Results: Student Academic performance

Regression Analysis

- Regression, correlation between scores:
  - GPA vs GRE Quant are positively correlated with a coefficient of 0.27
  - GPA vs GRE Verbal are positively correlated with a coefficient of 0.01
  - GPA vs GRE Writing are negatively correlated with a coefficient of -0.16

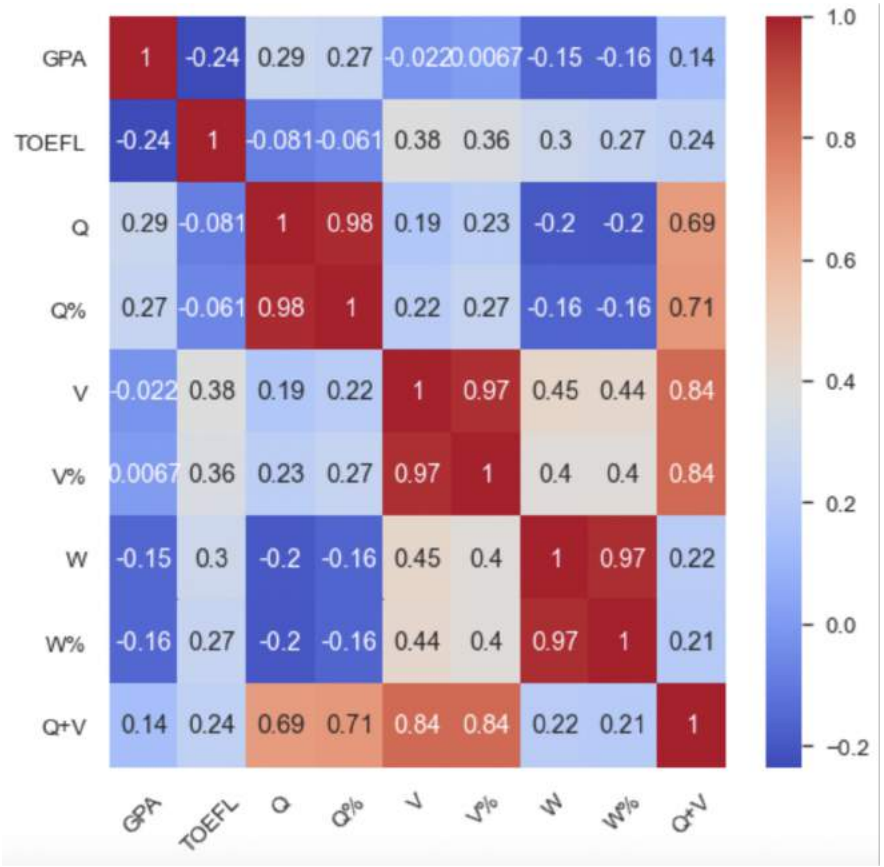






Results: Student Academic performance

Correlation Heatmap:

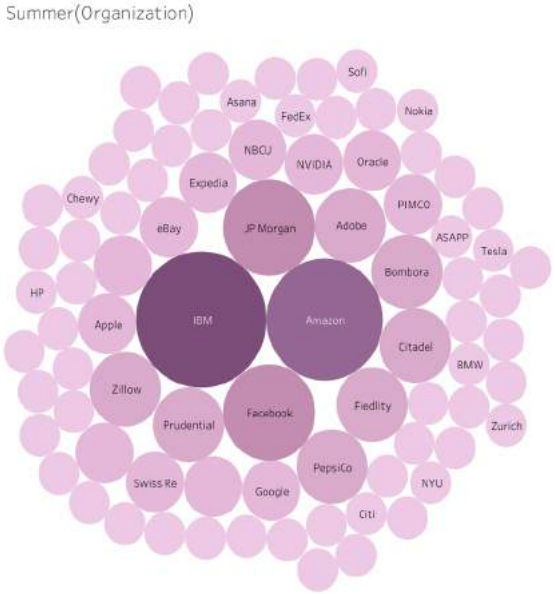


Results: Student Career Outcomes

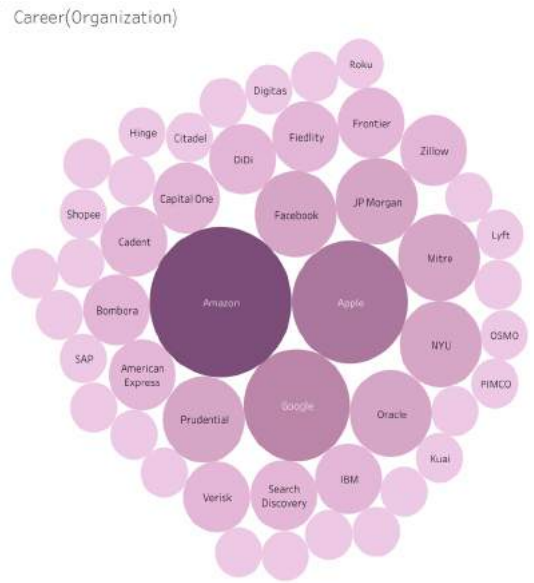
Bubble Charts of CDS Careers

Students pursued internships or full-time jobs in big tech companies such as Amazon, Apple, Google, IBM and Facebook

Summer Internship: Organization



Career Outcome: Organization

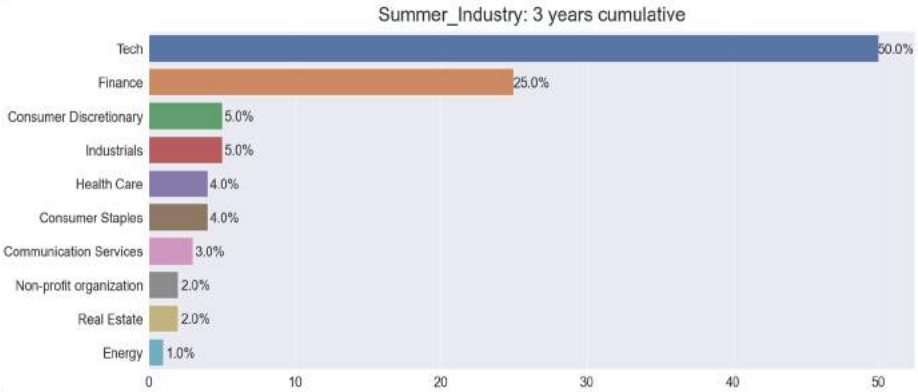


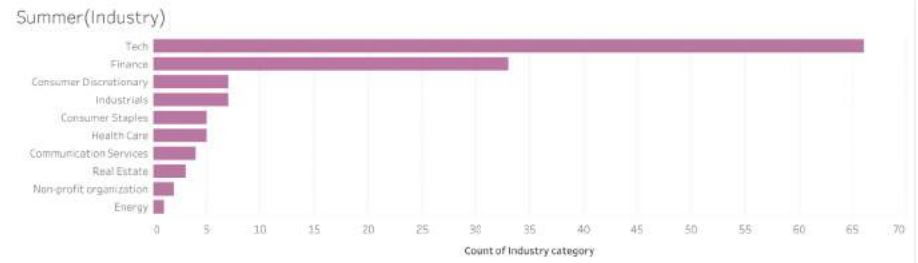
Results: Student Career Outcomes

CDS Careers by Industry visualized using matplotlib and tableau

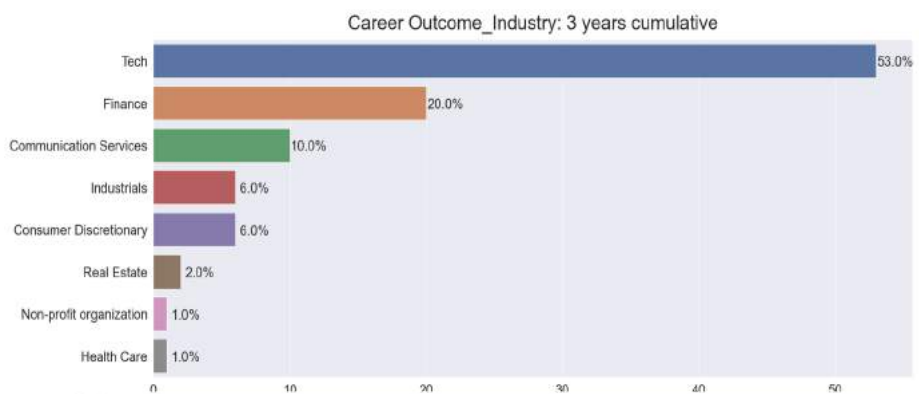
- Internships were focused in Tech (50%) and Finance (25%).
- Full time jobs were focused in Tech and Finance as well

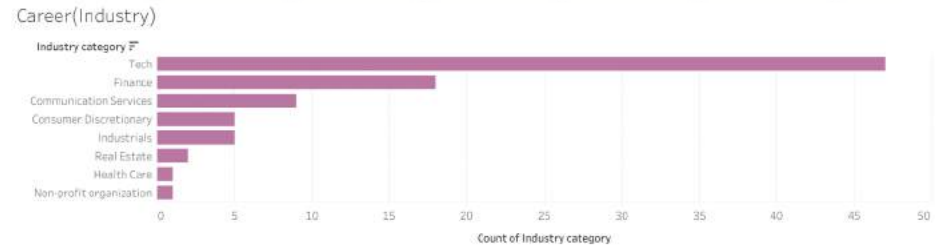
Summer Internship: Industry





Career Outcome: Industry





Results: Student Career Outcome

Company Size

- Most of the internships and full time offers were in large companies with more than 10,000 employees.

Summer Internship: Company Size

	Company_size
10,001+	0.59
1,001-5,000	0.14
51-200	0.08
11-50	0.06
5,001-10,000	0.04
2-10	0.04
501-1,000	0.04
201-500	0.02

Career Outcome: Company size

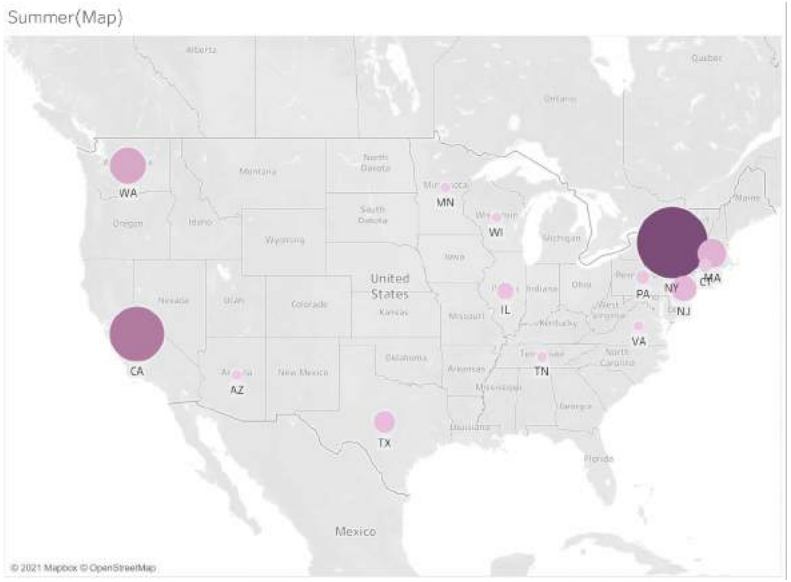
	Company_size
10,001+	0.65
5,001-10,000	0.10
51-200	0.09
1,001-5,000	0.08
11-50	0.03
201-500	0.03
501-1,000	0.01

Results: Student Career Outcomes

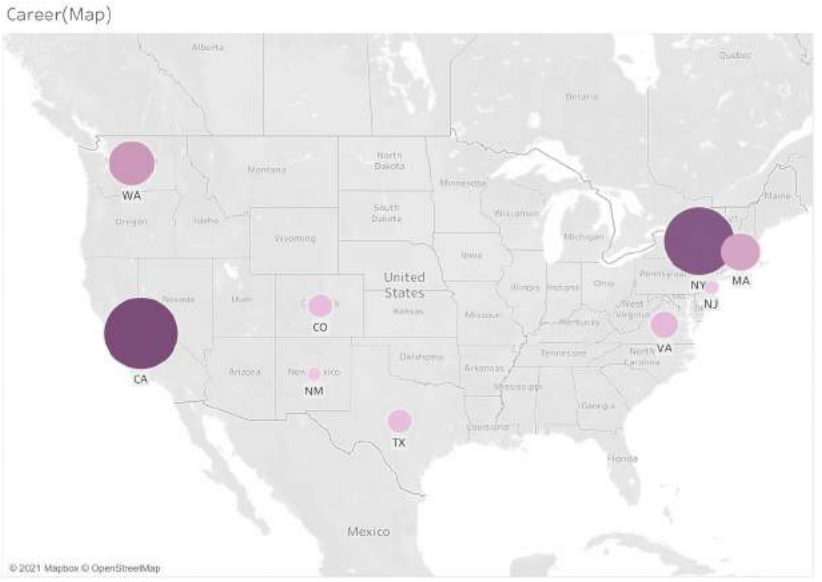
Company Locations

- Internships were primarily located at NY(41%), CA(24%), WA(11%), MA(7%). Location of Career Outcome distributed similarly

Summer Internship: Company Location



Career Outcome: Company Location

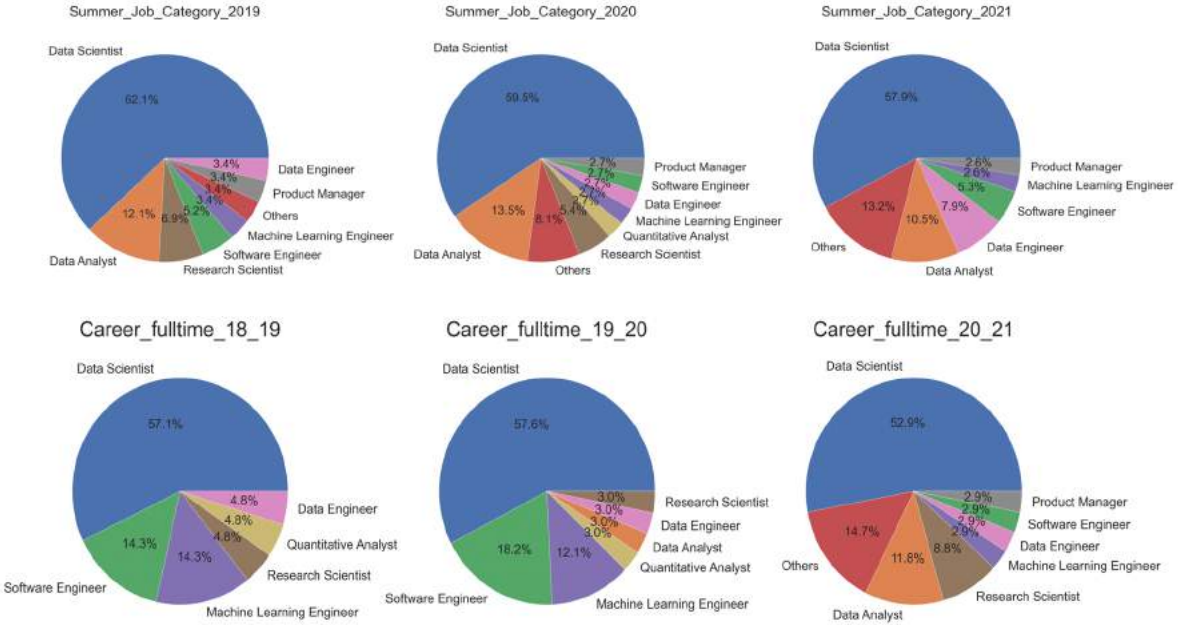


Results: Student Career Outcomes

Job Positions

- The majority of students are usually employed as data scientists or data analysts as interns. For full time offers, most students got roles as data scientists, followed by software engineers and machine learning engineers.

Summer Internship & Career Outcome: Job Category Trend



Sources:

- Lecture notes - numpy, pandas, plots and graphs: https://brightspace.nyu.edu
- Seaborn: https://seaborn.pydata.org/
- Matplotlib: https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.plot.html
- Tableau official website: https://www.tableau.com
- Global Industry Classification Standard: https://en.wikipedia.org/wiki/Global_Industry_Classification_Standard
- Gender composition in STEM:https://www.aauw.org/resources/research/the-stem-gap/
- Linkedin Company information: https://www.linkedin.com/jobs
- Current CDS webpage: https://cds.nyu.edu/careers-ds/
- CMU career outcome statistics: https://www.cmu.edu/career/outcomes/post-grad-dashboard.html
- MIT class profile: https://mitsloan.mit.edu/master-of-business-analytics/admissions/class-2022-profile