Best Majors for Financial Success

By: Dan, Dipika, Vikram

Agenda

- Problem statement
- Project planning and implementation
- Who did what
- Dataset description
- Dataset cleanup
- Pre Cluster graph
- K-means Clustering
- Cluster plots
- Libraries
- Challenges
- Lessons learned
- Summary

Problem Statement

Question
- Which majors provide the best monetary gain?

Data
- WSJ PayScale year-long survey of 1.2M people

Approach
- Perform dataset cleanup
- Apply unsupervised learning
    - Clustering using kmeans
- Generate PDF report to visualize results
    - Show cluster graphs

Project planning and implementation

- Assemble team fully committed to project
  - Gather 3-6 members for a group
  - Have each team member propose dataset for the project
  - Choose dataset by vote
  - Assign roles
- Setup and agree upon ground rules
  - Setup communications channels
    - Chat: whatsapp
    - AV: zoom
  - Setup repository: google drive
  - Plan minimal meeting time via 15' daily scrums at 6 PM CST Mon-Fri
    - Prerequisites
    - Active presence via daily (code & ppt) deliverables

Who did what

Dan
- Problem statement
- Project planning and implementation
- Web page harvest (new library)
- PDF report

Dipika
- Dataset cleanup
- K-means clustering of start vs mid career salary

Vikram
- Visualizations
  - Histograms and barplots

Dataset description

1	Undergraduate Major	Starting Median Salary	Mid-Career Median Salary	Percent change from Starting to Mid-Career Salary	Mid-Career 10th Percentile Salary	Mid-Career 25th Percentile Salary	Mid-Career 75th Percentile Salary	Mid-Career 90th Percentile Salary
2	Accounting	$46,000.00	$77,100.00	67.6	$42,200.00	$56,100.00	$108,000.00	$152,000.00
3	Aerospace Enginee	$57,700.00	$101,000.00	75	$64,300.00	$82,100.00	$127,000.00	$161,000.00
4	Agriculture	$42,600.00	$71,900.00	68.8	$36,300.00	$52,100.00	$96,300.00	$150,000.00
5	Anthropology	$36,800.00	$61,500.00	67.1	$33,800.00	$45,500.00	$89,300.00	$138,000.00
6	Architecture	$41,600.00	$76,800.00	84.6	$50,600.00	$62,200.00	$97,000.00	$136,000.00
7	Art History	$35,800.00	$64,900.00	81.3	$28,800.00	$42,200.00	$87,400.00	$125,000.00
8	Biology	$38,800.00	$64,800.00	67	$36,900.00	$47,400.00	$94,500.00	$135,000.00

- The dataset consists of 8 attributes and 50 records
- Each record is a undergrad major degree with its associated salary values

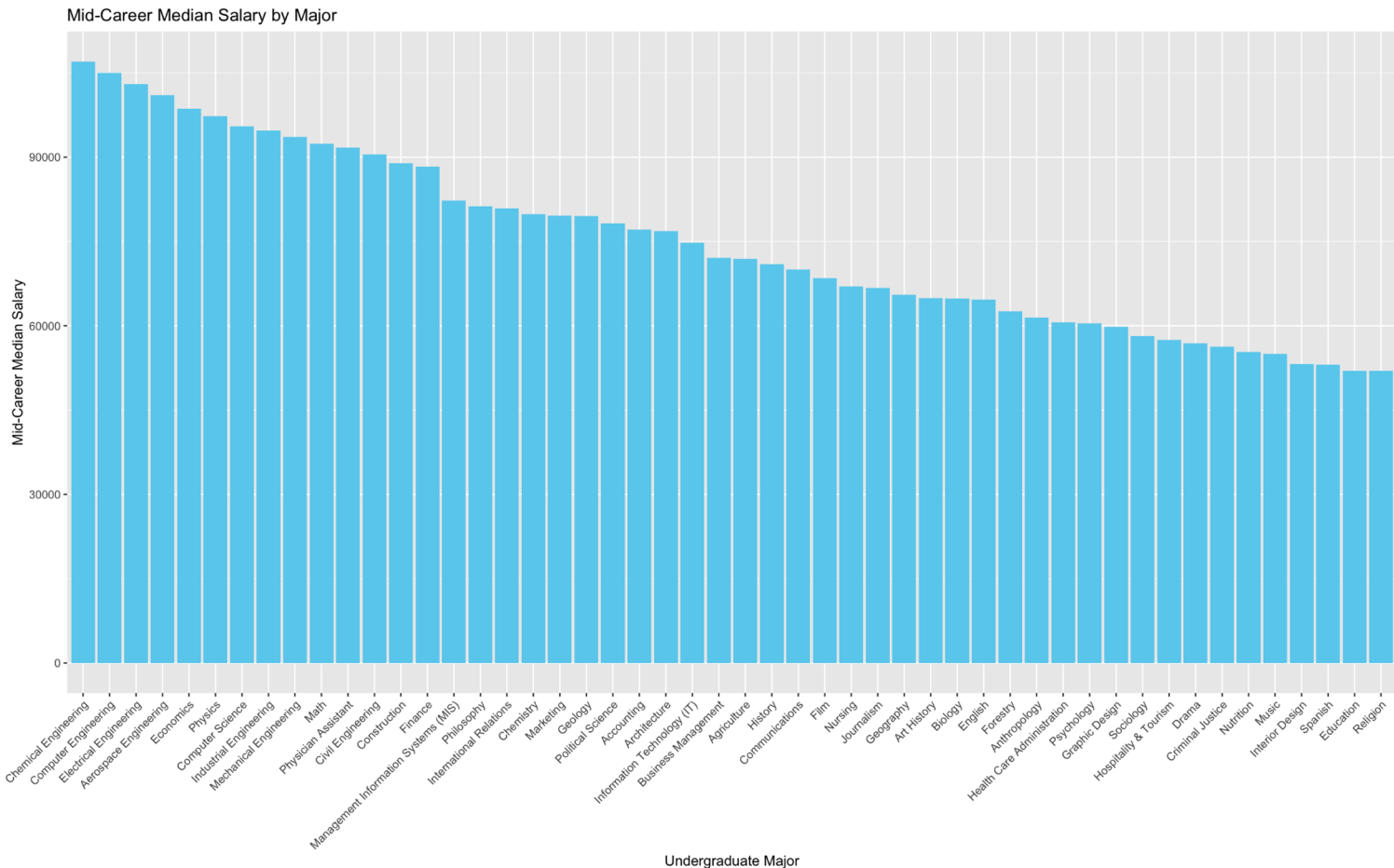# Dataset Cleanup

Salary values cleanup

1. Remove any dollar symbols
2. Omit commas in values
3. Convert to numeric values

Removal of a calculated column

1. Percent change from Starting to Mid-Career Salary was not required
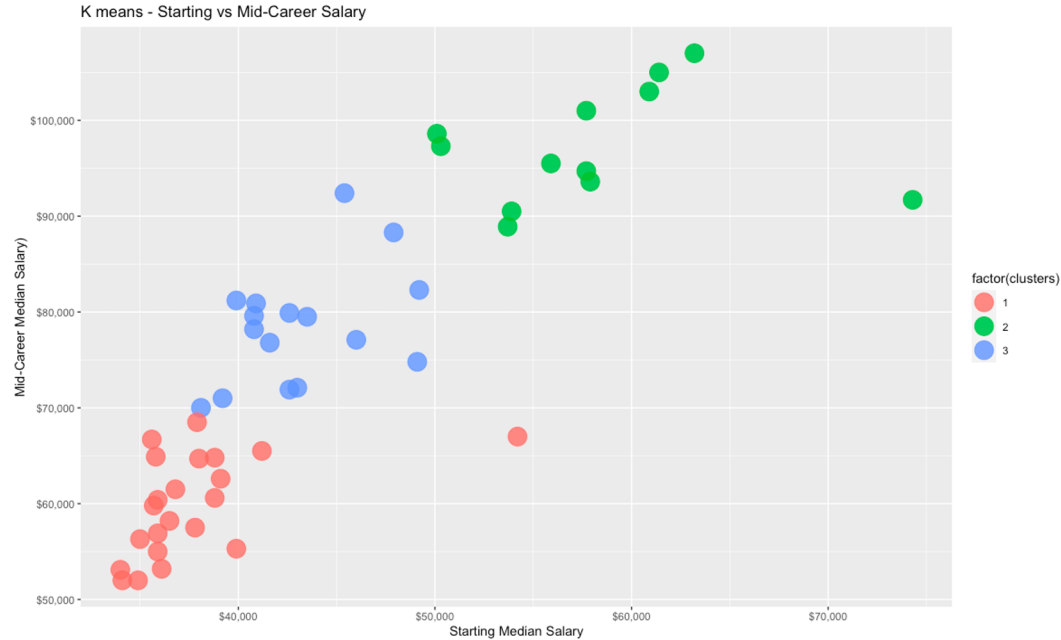   a. Can be calculated from Starting Median Salary and Mid-Career Median Salary
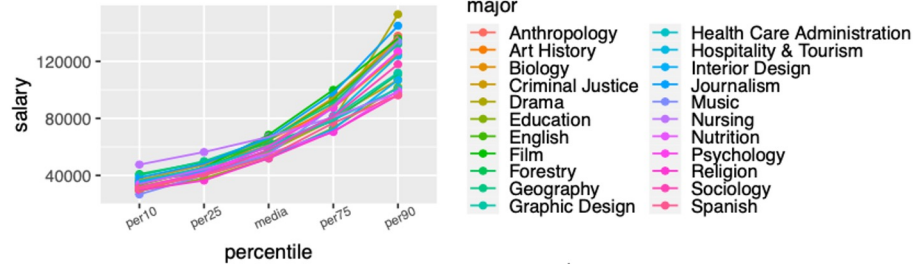
Pre Cluster Graph

Mid-Career Median Salary by Major

K-Means Clustering of Start vs Mid-Career salary

K means - Starting vs Mid-Career Salary

Undergrad degrees in the green cluster have the best pay back in terms of salary. They include:

1. Aerospace Engineering
2. Chemical Engineering
3. Computer Engineering
4. Electrical Engineering
5. Industrial Engineering
6. Mechanical Engineering
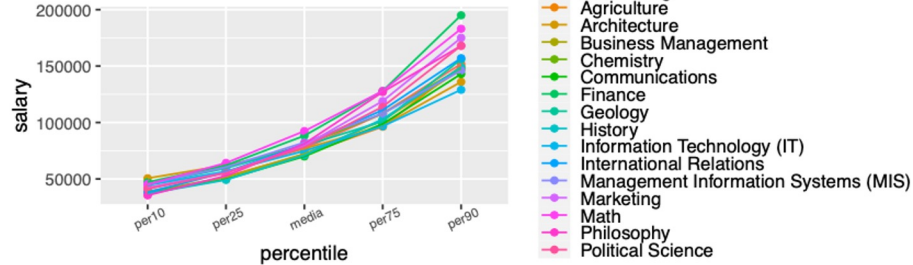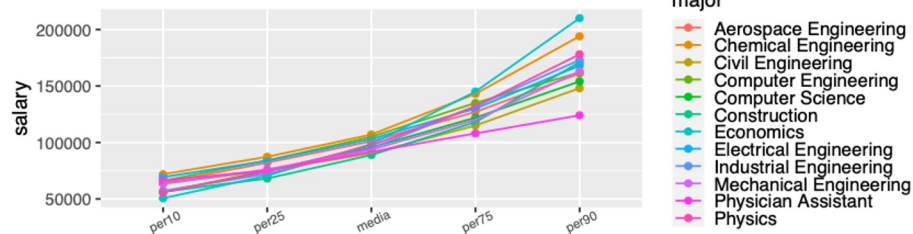
Cluster plots

Libraries

- base
  - as.numeric, gsub, scale
- dplyr
  - mutate_at
- factoextra
  - fviz_cluster, fviz_nbclust
- ggplot2
  - ggplot

- graphics
  - barplot, boxplot, hist, plot
- gridExtra
  - grid.arrange
- rvest
  - html_nodes, html_table, read_html
- stats
  - kmeans

Challenges

- Project Related
    - Dipika: Search for relevant functions and importing them from libraries
    - Vikram: Dataset changes for barplots and relevancy for said plots
- Teamwork
    - Dan: assemble a group fully committed to the project
    - Dipika: Cope up with daily scrums and deliverables
    - Vikram: Time conflicts for scrum meetings, creating zoom links

Lessons Learned

- Project related
    - Write clean code with comments - for easy readability and understanding.
    - Test code snippet if it works before appending to main script. Then, run the entire code to check for errors.
    - Analyze the latest script update on the version control before overwriting the script.
- Teamwork
    - Assign tasks to members as per their skills/strengths.
    - Discuss problems faced with other team members to get inputs to solve the issue as soon as possible.
    - Give constructive feedback on work contributed by other members.
    - Done through daily scrum meetings & whatsapp

Summary

- Highest Paying Degrees
  - STEM
    - $60K - 100K+
- Median Pay Degrees
  - Business
    - $45K - 60K
- Lowest Paying Degrees
  - Arts
    - $34K - 44K
- Live Demo