Writeup - Earnings Based On College Majors

Objectives

Below is the list of questions that we are looking to answer as a final outcome of this project:

- How much the college grads from different majors are earning after completing their education?

Goal Significance

Why do we need the earning data for college majors? What benefit we could derive by visualizing the earning data comparison between various college major grads? Below are the goals we can enlist:

- This information will give us an idea about the current market trends: i.e.
  - Understand the types of skill set are in demand
  - What are the extent of demand for various skills in general and in combination thereof under various fields.
- This will help us to structure our future college course structure that can fulfil the industry requirements and help develop best placement opportunities for our future students.
- This inference would also help us to frame our long-term & short-term goals and align the research activities with the industry demand/trends.

Data

Data Source

The original data on job was released by American Community Survey, which conducts surveys and aggregates the data and available at:
https://www.census.gov/programs-surveys/acs/

The data set cleaned by FiveThirtyEight is available for download at:
https://github.com/fivethirtyeight/data/tree/master/college-majors

Data Lists

The data is available in the .csv file format. File name: `recent-grads.csv`

Data Sampling Methods

Data Extraction Details

The dataset shows different major in college in every row that contains information on student gender diversity, their employment rates, overall median salaries, etc. Some of the columns that carry info of our interest are:

Rank - Rank by median earnings (the dataset is ordered by this column).
Major_code - Major code.
Major - Major description.
Major_category - Category of major.

Writeup - Earnings Based On College Majors

Total - Total number of people with major.

Sample_size - Sample size (unweighted) of full-time.

Men - Male graduates.

Women - Female graduates.

ShareWomen - Women as share of total.

Employed - Number employed.

Median - Median salary of full-time, year-round workers.

Low_wage_jobs - Number in low-wage service jobs.

Full_time - Number employed 35 hours or more.

Part_time - Number employed less than 35 hours.

Model

Highlights of the code.

Software packages used:

- Python
- Pandas
- Numpy
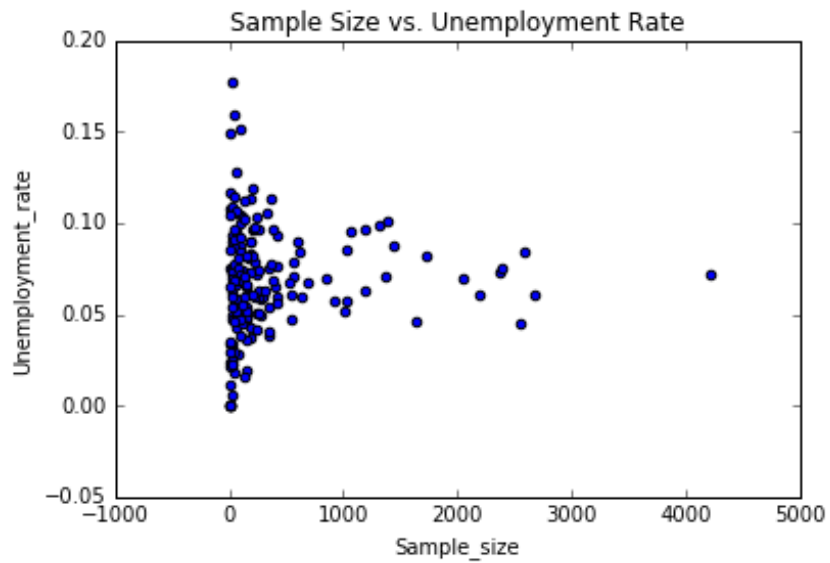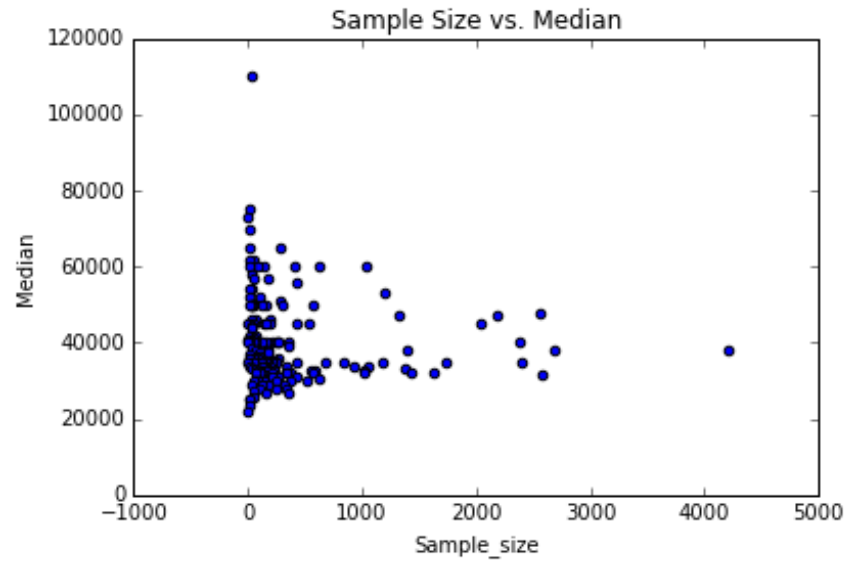- Matplotlib.pyplot

Overview:

- Read the data and form the data frame
- Identify the relevant info and segregate it
- Plot scatter diagram to understand their correlations
- Plot histogram to understand the data distribution
- Plot scatter matrix to visualize the overall picture of data dependencies
- Generate bar charts to review and interpret the info
- Prepare box plot to get the data statistics
- Create Hexagonal Bin Plot to visualize data density
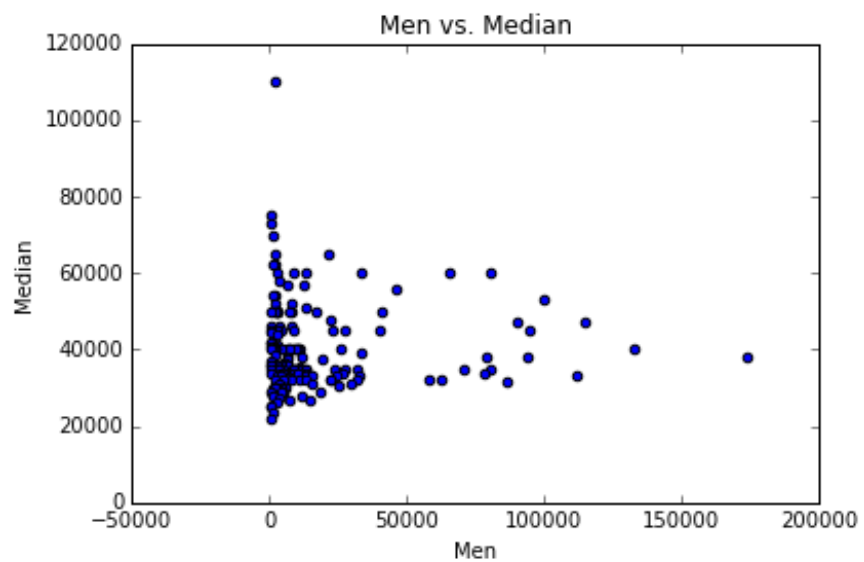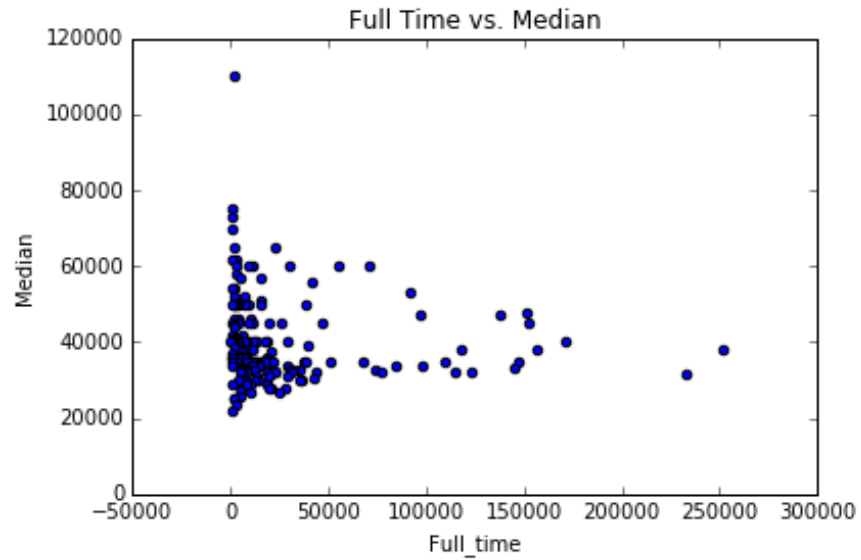
Writeup - Earnings Based On College Majors
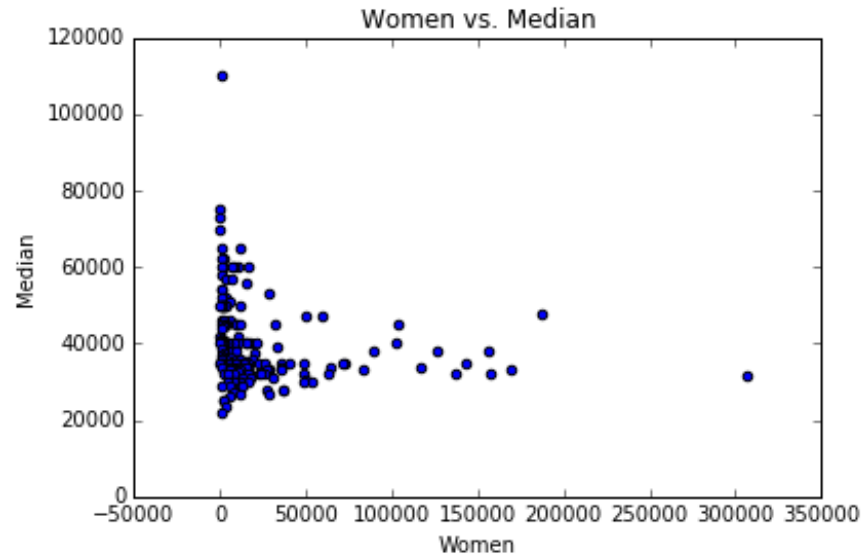
Results
Visualize the results.
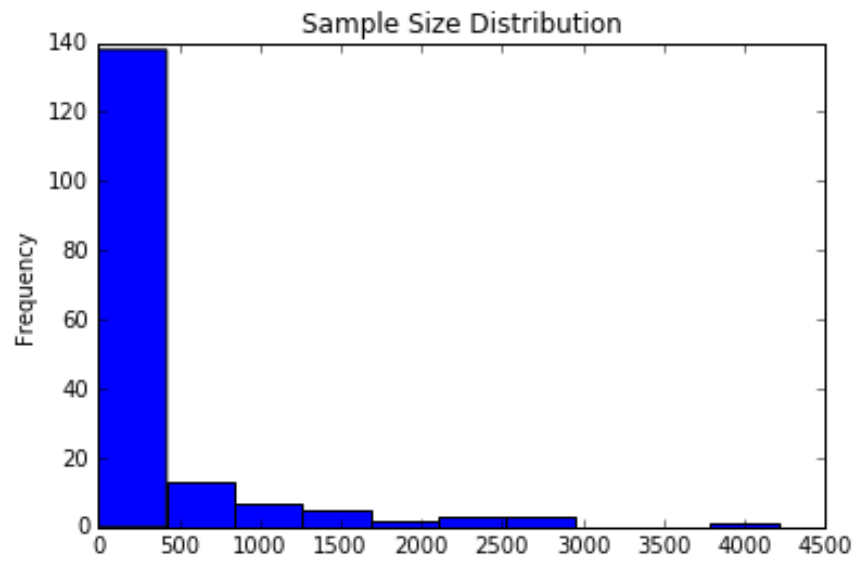
- **Scatter Plots to Review Data Correlations**

Writeup - Earnings Based On College Majors

Full Time vs. Median
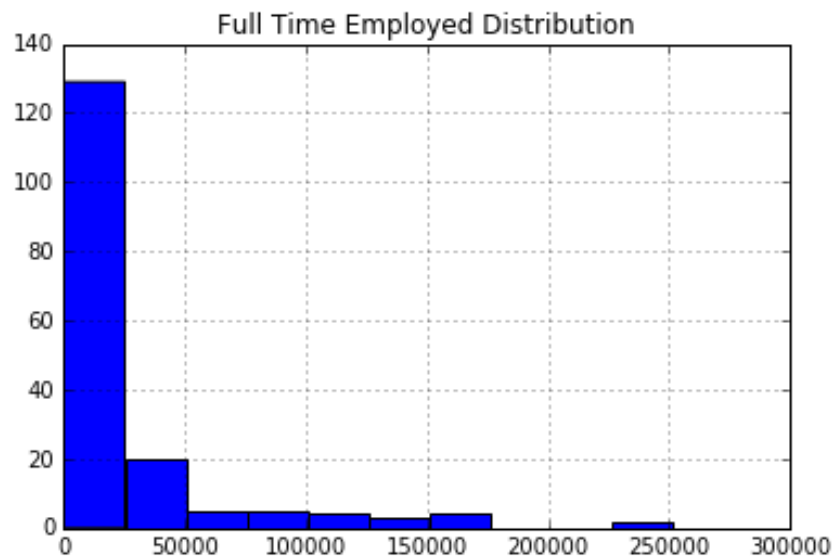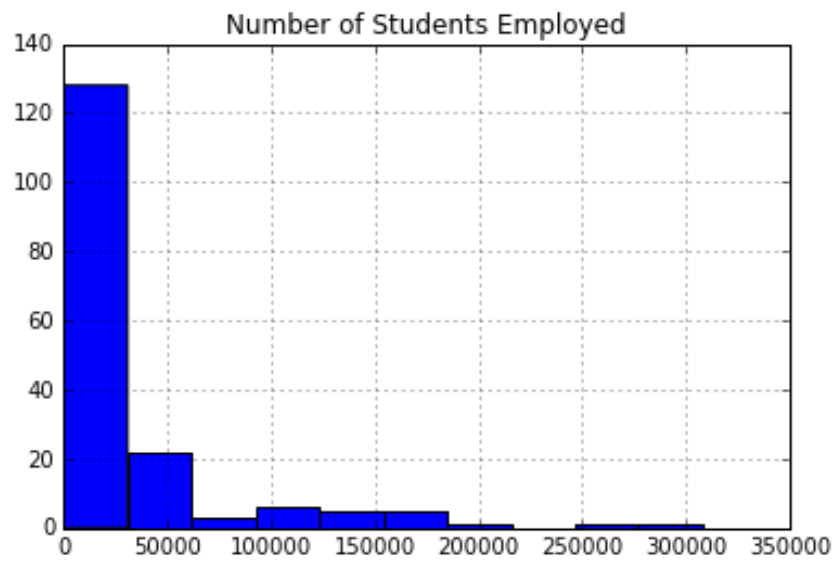

ShareWomen vs. Unemployment Rate

Men vs. Median

Writeup - Earnings Based On College Majors

Women vs. Median

- **Histograms to Verify Data Distribution**

Sample Size Distribution

Writeup - Earnings Based On College Majors

Median Distribution

Number of Students Employed

Full Time Employed Distribution

Writeup - Earnings Based On College Majors

Proportion of Women Share

Unemployment Rate Distribution

Writeup - Earnings Based On College Majors

Male Gender Distribution

Female Gender Distribution

Writeup - Earnings Based On College Majors

- **Scatter Matrix Plot for Data Statistics**
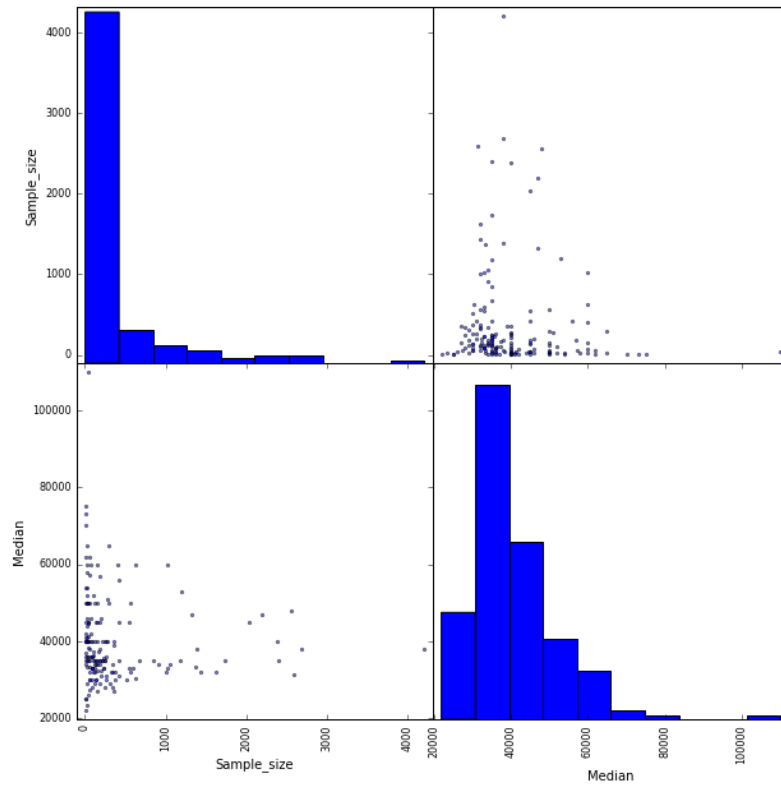
- **Scatter Matrix Plot for Unemployment Rate**
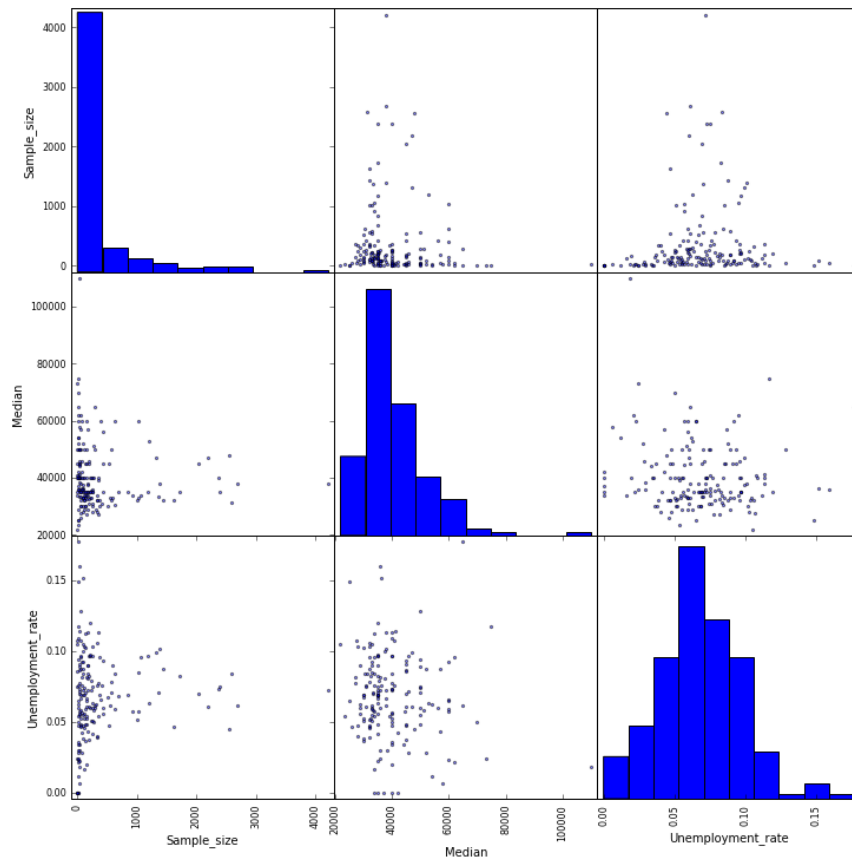
Writeup - Earnings Based On College Majors

- **Women Proportions in Different Majors**

Proportion of Women in Top 10% Paying Majors

Proportion of Women in Bottom 10% Paying Majors
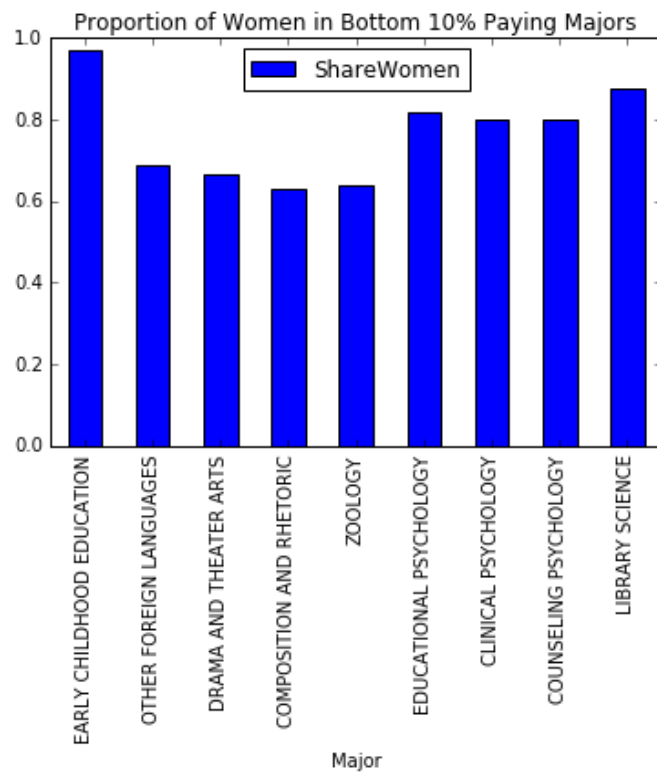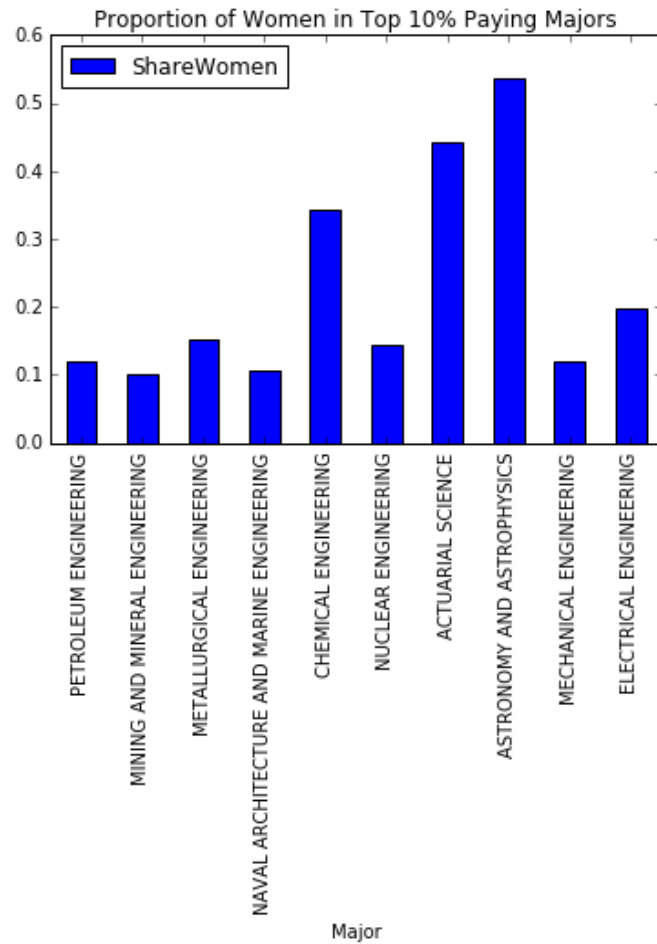
Writeup - Earnings Based On College Majors

- **Unemployment Rates in Different Majors**

% of Unemployment Rate in Top 10% Paying Majors

% of Unemployment Rate in Bottom 10% Paying Majors
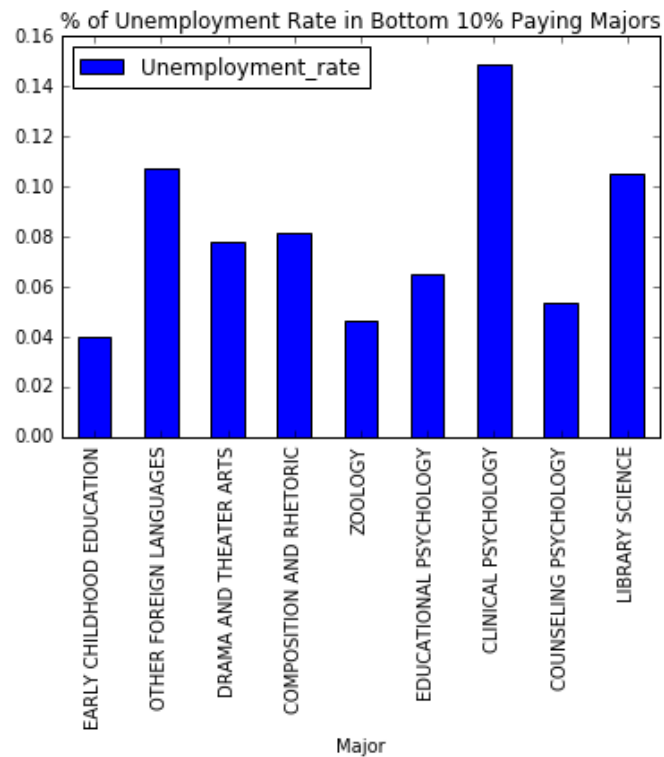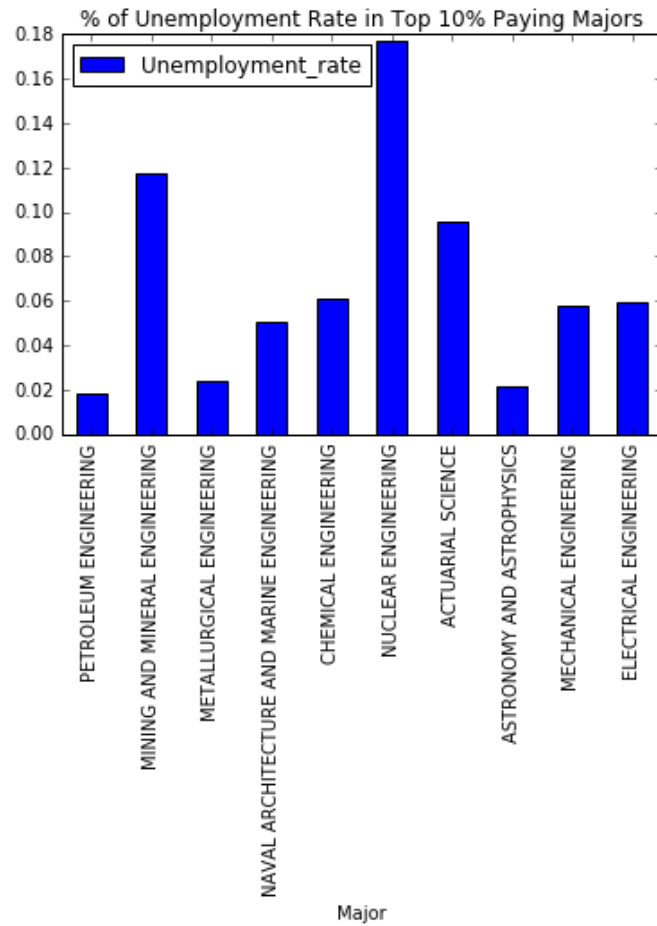
Writeup - Earnings Based On College Majors

- **Number of Men & Women in Each Major Category**
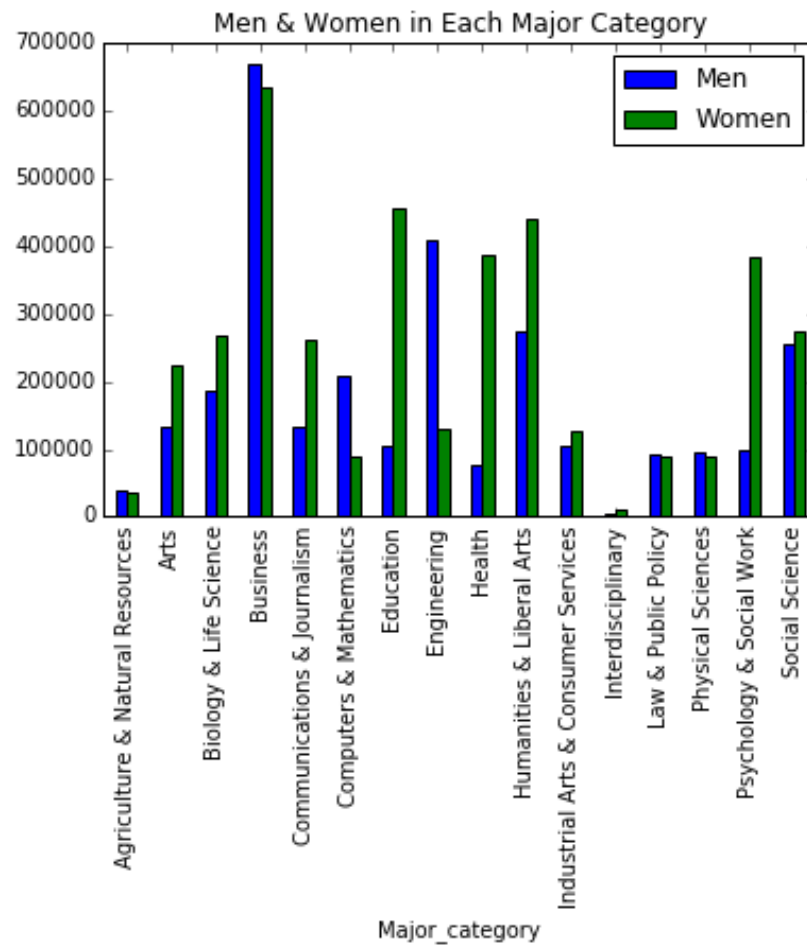
- **Distribution of Median Salaries & Unemployment Rate**

Writeup - Earnings Based On College Majors

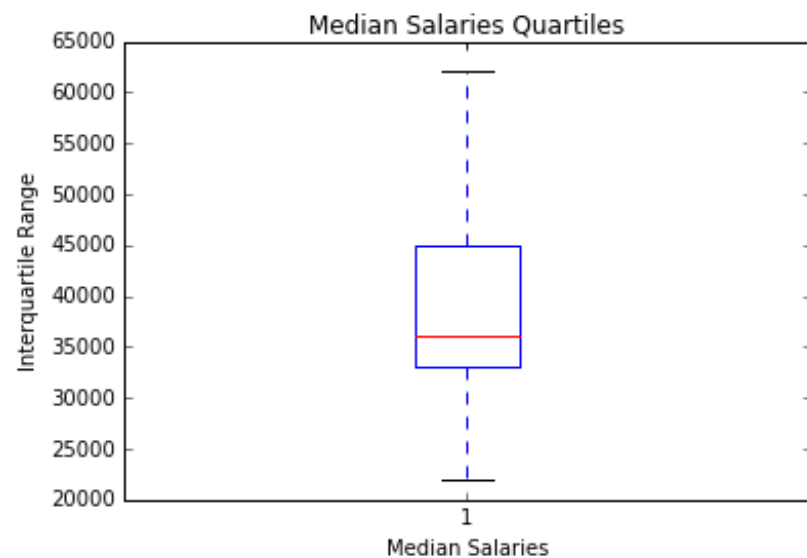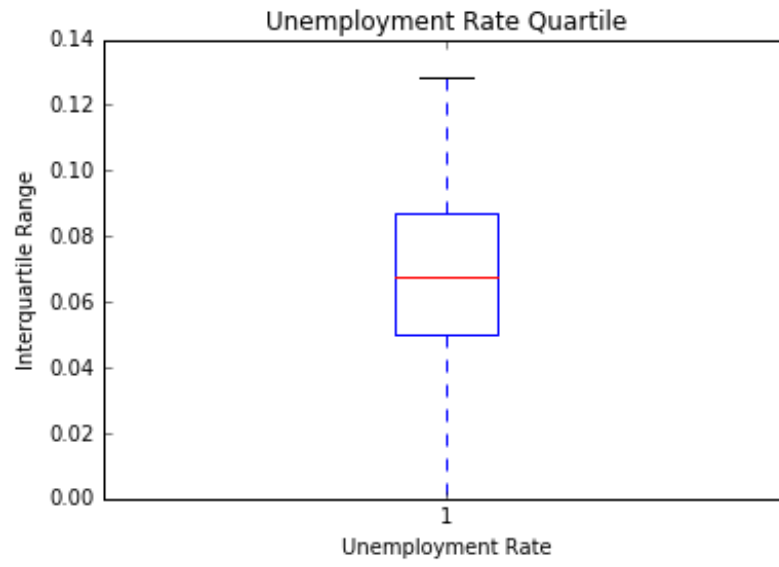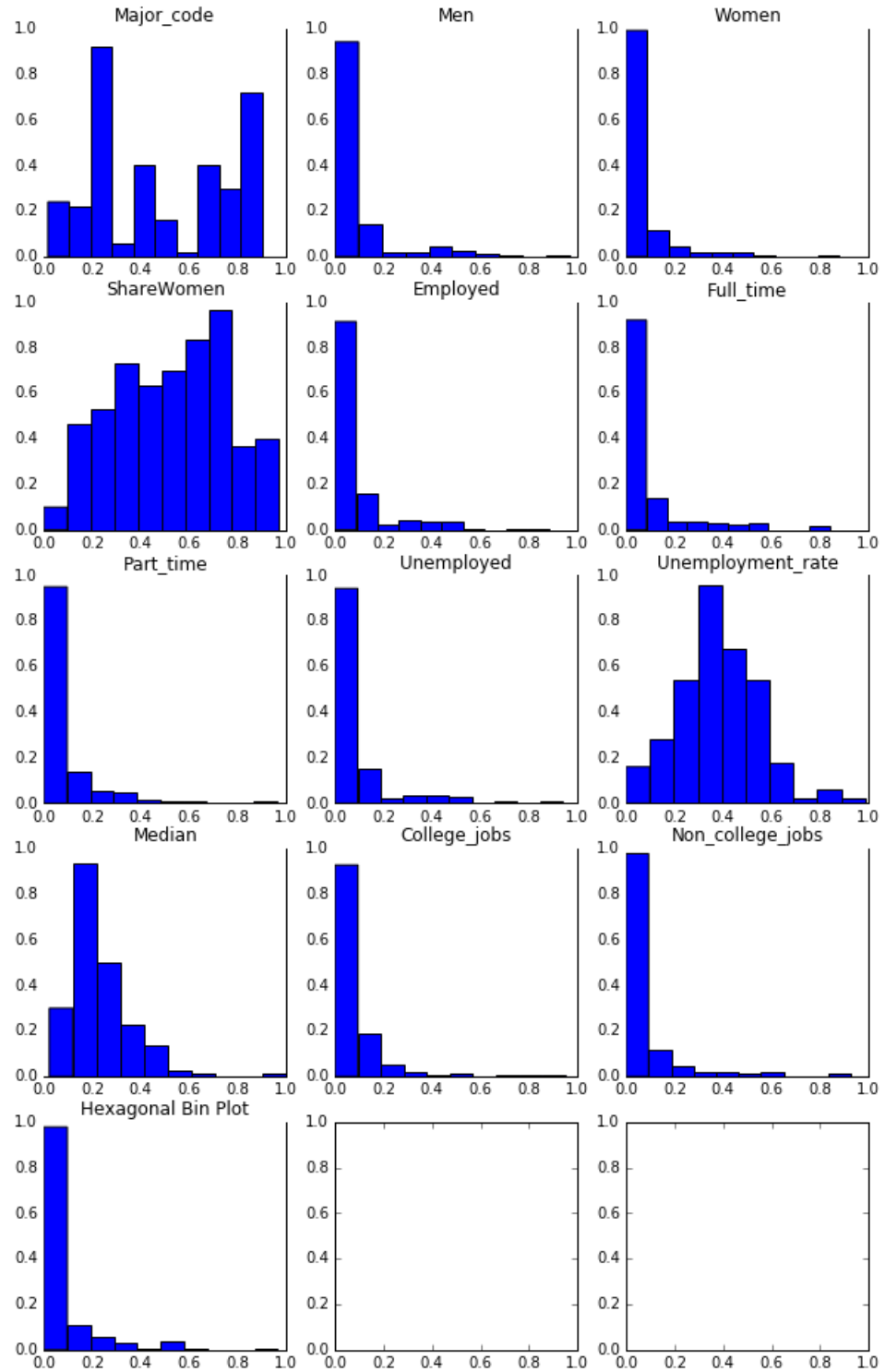# Writeup - Earnings Based On College Majors

- **Hexagonal Bin Plot to Visualize Data Density**

Writeup - Earnings Based On College Majors

Explain the results in a simple, concise and easy way. (non-technical audience)

The analysis and results give very useful info to shape our future planning based for:

- Couse mix
- Curriculum based on Industry trends
- Corporate tie-up for the skill set in demand
- Marketing and segmentations

Explain the results in the most technical way. (technical, data-scientist audience)

- Strong correlation has been observed between:
    - Sample Size vs. Median (of Earnings)
    - Sample Size Vs. Unemployment Rate
    - Full Time Employment Vs. Median (of Earnings)
    - Men (Gender Specific) Vs. Median (of Earnings)
    - Women (Gender Specific) Vs. Median (of Earnings)

- No correlation has been observed between:
    - Proportion of Women Vs. Unemployment Rate
    - Unemployment Rate Vs. Median (of Earnings)

- Data distribution results:
    - Most of the samples are found of size up to 500.
    - The median annual salary range appears to be 30,000 – 40,000
    - The distribution of full-time employed grads is found highly skewed across the course majors.
    - Distribution of the woman proportion across the course majors appears to be uneven, i.e. not normally distributed.
    - Unemployment rate is found following close to the normal distribution across the course majors.
    - Gender distribution across the course majors is found highly skewed, indicating a strong gender specific inclination towards specific course majors.

- For women,
    - Following are the top 10% paying majors:
        - Petroleum Engineering, Mining and Mineral Engineering, Metallurgical Engineering, Naval Architecture and Marine Engineering, Chemical Engineering, Nuclear Engineering, Actuarial Science, Astronomy and Astrophysics, Mechanical Engineering, and Electrical Engineering.
    - Following are the bottom 10% paying majors:
        - Library Science, Counselling Psychology, Clinical Psychology, Educational Psychology, Zoology, Composition and Rhetoric,

Drama and Theatres, Other Foreign Languages, and Childhood Education.

- The unemployment rate is found:
    - Higher amongst the following top 10% paying majors:
        - Nuclear Engineering
        - Mining and Mineral Engineering
        - Actuarial Science
    - Lowest amongst the following top 10% paying majors:
        - Petroleum Engineering
        - Astronomy and Astrophysics
        - Metallurgical Engineering
    - Highest amongst the following bottom 10% majors:
        - Clinical Psychology
        - Foreign Languages
        - Library Science
    - Lowest amongst the following bottom 10% majors:
        - Childhood Education
        - Zoology
        - Counselling Psychology

- Gender distribution amongst the major categories:
    - Both, Men and Women are equally inclined towards the following majors:
        - Business
        - Social Science
        - Physical Sciences
        - Low & Public Policy
    - Women prefers following majors to men:
        - Education
        - Humanity & Library Arts
        - Health
        - Psychology & Social Work
        - Communications and Journalism
    - Men prefers following majors to women:
        - Engineering
        - Computers & Mathematics
- Median of Earnings across the majors are as follows: (all numbers are approx.)
    - 25% majors are found having median earnings of up to $32,500
    - 25% majors are found having median earnings of between $32,500 - $36,000
    - 25% majors are found having median earnings of between $36,000 - $45,000
    - 25% majors are found having median earnings of above $45,000

Writeup - Earnings Based On College Majors

Conclusion

What we learn from this outcome. (non-technical audience)

- By this study, our understanding of the upcoming industry trends and change in the demand for the skill sets got improved. This would help us to shape our long-term courses proposition and industry tie-up to align with the market demand.
- This analysis can also help us to tune up our course majors based on the high and low unemployment rates in different sectors.
- The results can be effectively used to plan and emphasize our marketing efforts in appropriate sectors, i.e. we can vigorously focus our marketing over the highly paying majors with low unemployment rate while carry out our re-structure efforts for low-paying sectors with high unemployment rates.
- The gender specific inclinations for certain majors can also help us to decide about our message copy for different market clusters.

Technical significance of the results. (technical, data-science audience)

- We learnt correlations between various factors that affect the earnings of grads from different majors.
- We came to know the distribution of various features amongst the clusters such as median salary, full-time employment, unemployment rate, gender participations in various majors etc.
- We also understand what the high paying and low paying majors for women are.
- There exist gender specific preferences for different majors: we realized what majors men prefers to women and vice versa.
- We realized what are the top 10% and bottom 10% paying majors wherein the unemployment rate was the highest and the lowest.
- We got info about the Earning quartiles.

Suggestion for Further Development

How differently you would have done this project, if get a chance to do it again. Why?

In case of gotten second chance to do this project again, I would have:

- Touched upon some more features such as,
  - College jobs,
  - Non-College Jobs
  - 25th & 75th Percentile
  - Part Time Employed, etc.
  to develop correlation matrix and determine their influence over the earnings for different course majors.
- Verified the sufficiency of the sample size and come up with the conclusive suggestion about:
  - the data sufficiency
  - need for more data
  - need for more features to be included in the survey, etc.

Writeup - Earnings Based On College Majors

- Tried to explore more analytical info that could help to make long-term strategic decisions.

Your suggestions for someone if s/he wants to further continue this work.

Someone could pick one or more of the untouched data fields, explore and add additional features and continue this journey further to cherish more flavors in the analysis. S/he could follow some of the suggestions that are mentioned above in the second chance effort.