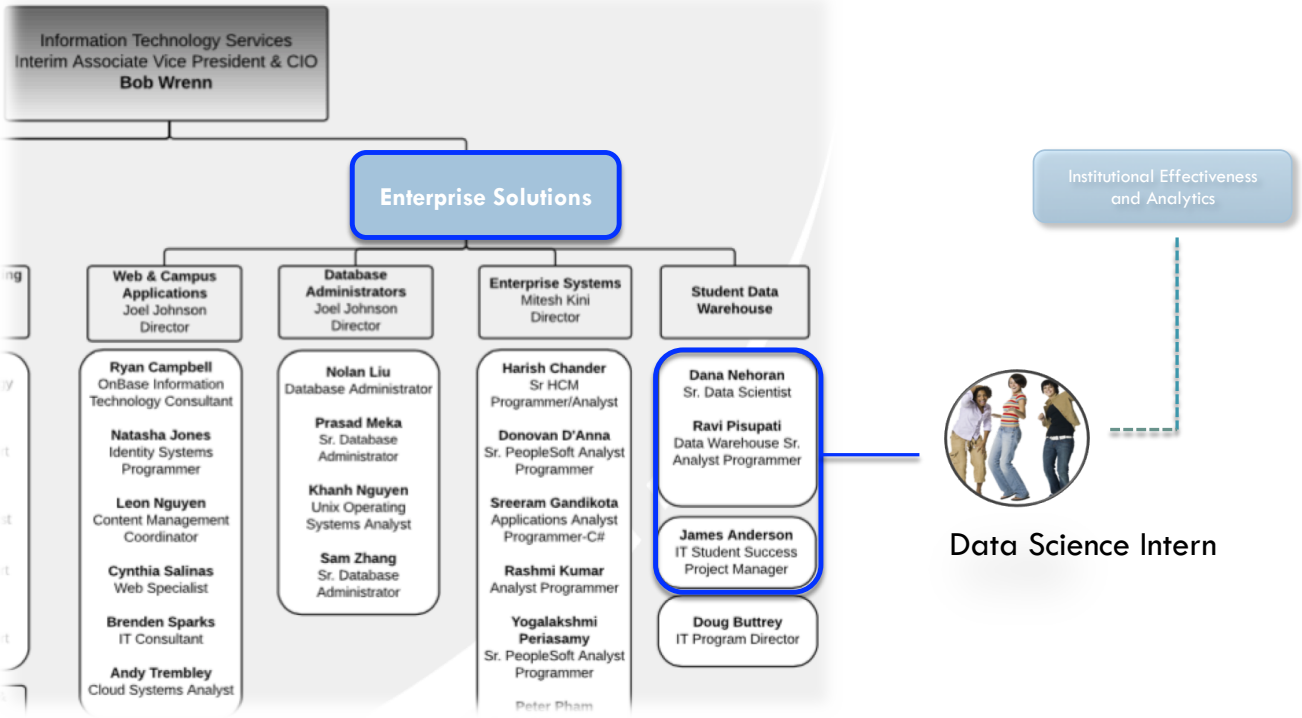
STATISTICS INTERNSHIP REPORT

Fengmei Liu
02/20/2017

Organization and Job Introduction

- SJSU ITS Enterprise Solutions - Data Science Intern

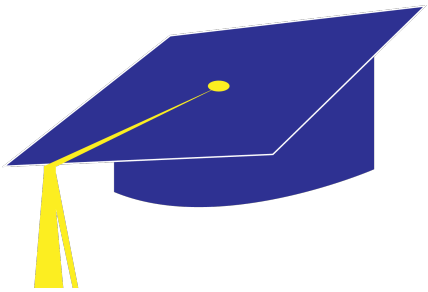


Internship Search Process

- This opportunity was from **Dr. Bremer's email** in the beginning of 2016 fall semester.
- **Being active, and relative class experience** prepared me in finding this internship.
- Suggestion: **"job search is a job"**

Work Project 1

- Analysis of unexpected graduate students in MIS major
  - Data extraction
    - The numbers: yearly graduation, admitted etc.
    - The information: where are they from, transfer schools, majors
  - Data visualization
  - Descriptive statistics
    - mean , median, variance, quintiles



Work Project 2

- Predictive model prototyping and building about students enrollment decision.
  - Data processing and missing data, categorical data analysis, imbalanced data processing.
  - Classification models- logistic regression, decision trees, Adaboost, gradient boosting machine.
  - Cross-validation and prediction, confusion matrix, ROC curve
  - Mixed data clustering, factor analysis

Software and Tools

- **R**
- Being fluent on different R packges for data visualization, like ggplot2, plotly etc.. And machine learning packages
- **SQL**: undergraduate class, learned on the job
- **Flexdashborad**: markdown, learned on the job
- **GitLab**

Skills to be successful during internship

- Communication
  - Work update and summary
  - Presentation
- Programming
- Analytical
  - Statistics
  - Problem solving

Contact Information

- Dana Nehoran (Sr. Data Scientist and Business School professor)
- This position is continuously hiring in the beginning of every semester
  - By referral
  - Dr. Bremer's email