CSC-450, Senior Research Project Ideas

Your *research* project must be centered around a specific question you will answer, a hypothesis that you will test, or an objective that you will meet by designing a method or model, collecting and analyzing data, and drawing conclusions using an appropriate and reproducible methodology.

The following is a list of possible research questions / projects, and is designed to help generate ideas. There are many, many more ideas that would be acceptable. Don't hesitate to contact me if you want to run any ideas by me.

Topic Area	Question / Hypothesis / Objective	Method
Data Mining*	What region of the U.S. is most interested in *cats*?	Based on Twitter data, compare the proportion of tweets that mention *cats* in the west, southwest, midwest, southeast, and northeast.
	Who are the highest-grossing movie directors? (https://www.r-bloggers.com/imdb-movie-analysis/)	Analyze IMDB movie data to summarize movie revenue by director.
	Any individual can be connected to any other individual in at most 6 steps (see six degrees of separation; six degrees of Kevin Bacon).	Using (a subset of)Twitter data (or Linked In data), determine the average number of connections separating any two individuals.
	What computer programming languages are employers looking for?	Write a script (better yet, create a web page) that downloads current programming job listings from a website such as Indeed.com, and tallies results for a set of languages
Database Design	A relational database (e.g., MySQL) is more efficient (allows for faster querying) than a NoSQL database (e.g., MongoDB) for a particular application.	Create both a MySQL and MongoDB database for the same underlying data, and compare the time it takes for various queries.
Digital Logic	Two's complement is a better (more efficient) way of representing signed numbers than one's complement	Implement each representation, and mathematical operations (e.g., addition), and compare the efficiency of mathematical operations on a series of numbers.

Computer Security	A password stored using *md5* encryption is easier to crack than a password stored using *bcrypt* encryption. (inspired by: https://blog.avast.com/2015/09/07/taking-a-closer-look-at-cracked-ashley-madison-passwords/)	Choose a set of passwords to analyze. Use a password cracking tool and evaluate how long it takes to crack each password. How does the time change with the complexity of the password?
	Method A can be used to detect spam e-mails (or spam tweets) (and is better than Method B)	Develop a method (e.g., a machine learning algorithm) to detect whether an e-mail is spam or not. Calculate the accuracy of the method across 500 e-mail messages. Compare the accuracy of this method to the accuracy of another method.
Modeling / Simulation	Placement of a pillar that partially obstructs a door results in more people escaping safely in the event of a fire.	Use an agent-based model to simulate individuals inside a room, that either has or does not have a pillar. At a fixed time, a fire is initiated, and the individuals run out of the room. Compare the speed and number of people safely exiting the room across the two conditions.
Game Development	There are currently no (good) games for students to learn (or reinforce) Java concepts	Create a fun, engaging game where students get points for demonstrating knowledge of Java concepts. (Ideally, you would evaluate whether students learned more (or had greater enjoyment) from this game than others).
	Which method is better for rendering objects?	Compare performance and efficiency between different methods for rendering objects
Any field	You may answer a question by conducting a survey, for example to evaluate how knowledgeable individuals are about security issues.	This research falls under Human Subjects Research (see the *HumanSubjects.pdf* file for more information)

*Many datasets can be found https://www.kaggle.com/datasets and http://rs.io/100-interesting-data-sets-for-statistics/