Predicting the Lottery

**Author Name:** Liam Rowell

**Professor:** Chris Kreider

**Executive Summary:**

The purpose of this project is to create a system to model university the housing lottery. This will reduce the uncertainty around housing lotteries and allow students to know what kind of housing will likely be available at their timeslot. The system will pull from housing data from google drive into a database, where it will then be processed. The model will then run and data will be made available through a web-app. The model, the database, the web-app and the processing tools will run on Amazon web services (AWS).

Contents

[Introduction 3](#_Toc63239906)

[Project Background 3](#_Toc63239907)

[Project Description 3](#_Toc63239908)

[Proposed Solution 3](#_Toc63239909)

[Development Approach 3](#_Toc63239910)

[High Level Plan 4](#_Toc63239911)

[Project Schedule 5](#_Toc63239912)

[Work Breakdown Structure 5](#_Toc63239913)

[Project Calendar 5](#_Toc63239914)

[Intermediate Milestones 5](#_Toc63239915)

[Weekly Updates 5](#_Toc63239916)

[Project Deliverables 6](#_Toc63239917)

[Conclusion 7](#_Toc63239918)

# Introduction

In the spring of every year, thousands of college students attempt to register for the housing they will use the following year. One of the more common ways to do this is to assign each student a time slot where they get to pick their housing. It is currently challenging to determine what housing will be available at any given time slot. The purpose of this project is to create a probabilistic model that will remove some of that challenge.

## Project Background

As a student at Christopher Newport University (CNU), I went through the housing lottery two times. I have friends who have gone through more. Every time was stressful because it was hard to determine how likely it would be for the most desired housing to be available at a given time. Last year, I watched it stretch the relationship between some of my friends who were living together. A probabilistic model would have let them know what kind of housing was likely to be available at their best time slot. They would have been able to make a more educated decision based on a defined risk tolerance rather than guessing at what would be available based on my back of the hand calculation at the time. If this is a problem at CNU it is likely an issue at other schools as well

## Project Description

The project should be capable of handling three different scenarios with slightly different inputs, which will always include a university and a time slot. The three scenarios are as follows:

1. The project should be able to determine the probability that a room at a given location and a given occupancy will be available at the given time slot.
2. The project should be able to determine the probability that a room in a given location will be available, and what the last occupancy available is most likely to be.
3. The project should be able to determine the probability that a room with a given occupancy and what the last location available is likely to be.

The project should be in the form a of a web application with visualization and explanation of the output. It is intended to be used by students with data provided by universities. The project may also provide information back to the universities on common student queries.

# Proposed Solution

## Development Approach

I intend to use the Agile Kanban, probably in the form of a Trello board. I choose agile over waterfall because I don’t want to lock myself in to specific program requirements. There is a good chance that once I get going new requirements and challenges may make themselves known. I choose Kanban as my approach because it is a very simple form of agile that works best with one developer. This will allow me to easily add and subtract tasks from different lists.

## High Level Plan

In terms of overall structure, the project will look like this:

Now, to break this very high level plan down into a slightly less high level plan. All the data for each university (or each lottery) will be stored in a google drive folder with its own defined structure. That data is then pulled and reformatted into a database, where it can be accessed by the model. The model is responsible for running the simulation and putting the created data in the database. The user can then interact with the web app, which will query the database and render the simulation. The system will use a variety of different languages and frameworks. Both the input tool and the model itself will be python programs running on an AWS EC2 instance. The input tool will leverage the google docs API to pull data from the google drive folders and MySQL connector to push it into the database. The database will be in MySQL and run on an AWS RDS. The front-end will be written in HTML and PHP and take advantage of bootstrap. It will run on an Apache server on the same EC2 instance as the Python programs.

The first phase of any project is analysis. I plan to create a set of user stories to flesh out exactly what the project needs to do. I then plan to create fully developed use cases as well as sequence diagrams from those sequence diagrams. I also want to create a layout for the project in balsamiq. After I have a clear idea of exactly what the project should do, I will create a database ERD that accommodates all the necessary data.

Each element of the project will be constructed separately and then connected as they are built. The first thing to be built will be the database. Building the database first sets a clear framework for how data should be sent in and out. Next, a skeleton of the front-end web app will be constructed. The main goal at this point in the project will be to make sure that all data can be displayed though not in the prettiest fashion. I am doing the front-end second because I am less comfortable with it than I am with many of the other aspects of the project. After that I plan to build the input tool, again with minimum functionality. Finally, I will build the model to be fully functional. At this point, I will have a fully functioning skeleton of the project and will begin to refine it, but in reverse order. I will adjust the input tool so that it can take a greater variety of inputs. Then I will refine the web-app. Finally, I will make any tweaks to the database that are necessary, however, I don’t expect that any tweaks to the database will be necessary.

At this point the project should be finished.

Fundamentally, my project has four major critical systems, the database, the input tool, the model, and the web app. Of these four, only the web-app is visible. I will build all four of these elements, but all four will use non-critical tools. I expect to have the most difficulty creating the web-app, as this is something that I have not done before. The model will probably be the second most difficult thing to build, because it is by far the most complex part of the project, which will also cause it to take the most time.

# Project Schedule

## Work Breakdown Structure

Bellow are the main tasks that need to be created for the project.

* Create User Stories (30 mins)
* Create fully developed use cases and sequence diagrams (1 hour)
* Create System Vision Document and outline stakeholders (1 hour)
* Design Database and create ERD (2 hours)
* Design GUI for webapp in balsamiq (1 hours)
* Implement Database on AWS RDS (2 hours)
* Set up EC2 with Apache and clone in git repository (30 mins)
* Code basic version of Web App and connect it to the database (3 hours)
* Create sample data (1 hour)
* Figure out how to use Google Drive API (1.5 hours)
* Code basic version of the input tool (1.5 hours)
* Create mathematical model for data (2 hours)
* Assess how to code the math in Python (1 hour)
* Code the database to model connection (1 hour)
* Code the math portions of the model (4 hours)
* Code the model to database connection (1 hour)
* Adjust the import tools to account for different ways to format the data (3 hours)
* Finish the design of the web-app so it looks nice (2 hours)
* Clean up, test, and prepare for presentation (4 hours)

## Project Calendar

### Intermediate Milestones

#### Intermediate Milestone #1

For intermediated milestone 1 I would like to present my analysis phase of the project, which should include my fully developed use cases and my ERD. I would also like to present a website which can contact my database (which should also be built but not filled at this point).

#### Intermediate Milestone #2

For intermediate milestone 2, everything except the actual mathematical model itself should be working. This means I should be able to show data going from google drive to the database to the model, and then back to the database again. If I am ahead of schedule the model will also be done, however, nothing will be as refined as the finished project.

### Weekly Updates

#### Weekly Update #1

* Create User Stories (30 mins)
* Create fully developed use cases and sequence diagrams (1 hour)
* Create System Vision Document and outline stakeholders (1 hour)
* Set up EC2 with Apache and clone in git repository (30 mins)

#### Weekly Update #2

* Design GUI for webapp in balsamiq (1 hours)
* Design Database and create ERD (2 hours)

#### Weekly Update #3

* Implement Database on AWS RDS (2 hours)
* Code Basic Web app (2 hours)

#### Weekly Update #4

* Create sample data (1 hour)
* Figure out how to use Google Drive API (1.5 hours)
* Code basic version of the input tool (1.5 hours)

#### Weekly Update #5

* Assess how to code the math in Python (1 hour)
* Code the database to model connection (1 hour)
* Code the model to database connection (1 hour)

#### Weekly Update #6

* Code the math portions of the model (4 hours)

#### Weekly Update #7

* Adjust the import tools to account for different ways to format the data (3 hours)

#### Weekly Update #8

* Finish the design of the web-app so it looks nice (3 hours)

#### Weekly Update #9

* Clean up, test, and prepare for presentation (4 hours)

# Project Deliverables

Documentation

* System Vision Document: brief description of the vision for this project
* Stakeholder Identification: Discussion of who the target users will be, and potential users in the future if relevant.
* User stories
* Fully Developed use cases and system sequence diagrams
* Database Entity Relationship Diagram for database
* GUI Mockups for web-app in balasamiq

Technical Deliverables

* Zip Archive with completed code for project and scripts for creating databases
* Link to URL with actively working project for instructor review
* Link to GIT repo with completed code for project including scripts for creating databases
* Sample data, as well as how a university could sign up to use the software and the process of implementation for a university.

*Final Presentation Deliverables*

* A full explanation of how a student would use the system.
* Full explanation of how a university would use the system.
* Explanation of how the systems interact with each other.

# Conclusion

The purpose of this project is to effectively model universities housing lotteries so that both students and staff have a more certain picture of what housing will be available at what time. The project should take about nine weeks. I am using new technical tools that I have not used before, which does present some risk, however, I have built time into my schedule to learn those new tools.