# 

# **CIS5560 Term Project Tutorial**

#### **Authors: Marco Aramburo, Henry Flores, Antonio Ponce, Michael Mijos**

#### **Instructor:** [**Jongwook Woo**](https://www.linkedin.com/in/jongwook-woo-7081a85)

#### **Date: 12/15/19**

**Lab Tutorial**

Marco Aramburo, Henry Flores, Antonio Ponce, and Michael Mijos

12/15/19

**Data Analysis using Microsoft Azure ML**

**Objectives**

**List what your objectives are.** In this hands-on lab, you will learn how to:

* Upload your dataset into Microsoft Azure ML
* Learn the different experiment items and what each do
* Fill in missing values
* Sort through the data columns and choose what we need
* Visualize the data

**Platform Spec**

· CPU Speed: 3.2ghz

· # of CPU cores: 6 cores

· Total Memory Size: 16gb

· Hard Drive: 350 gb

Step 1: Setting up your experiment

**Explain what this step is for.** This step is to get data manually….

Download the dataset via this link: <https://www.kaggle.com/chicago/chicago-food-inspections>

1. Click download and once it is finished go to your downloads folder. Extract the dataset.
2. In Microsoft Azure ML, accessible through this link. <https://studio.azureml.net>. Create an account or sign into it if you already have an account.
3. Once you are signed in, click the ‘+’ at the bottom left of the screen. Once it window is opened, click the three cylinders labeled ‘Datasets,’ and click the folder labeled, “from local file.’
4. Find and select the downloaded dataset and upload it.
5. Once it is uploaded, in the same ‘+’ menu, go to experiments and click blank experiment.
6. Name this experiment, “Chicago Food Inspections.”

Step 2: Implementing the modules

**Explain what this step is for.** In this step we are going to load in all the modules necessary in creating a good visual.

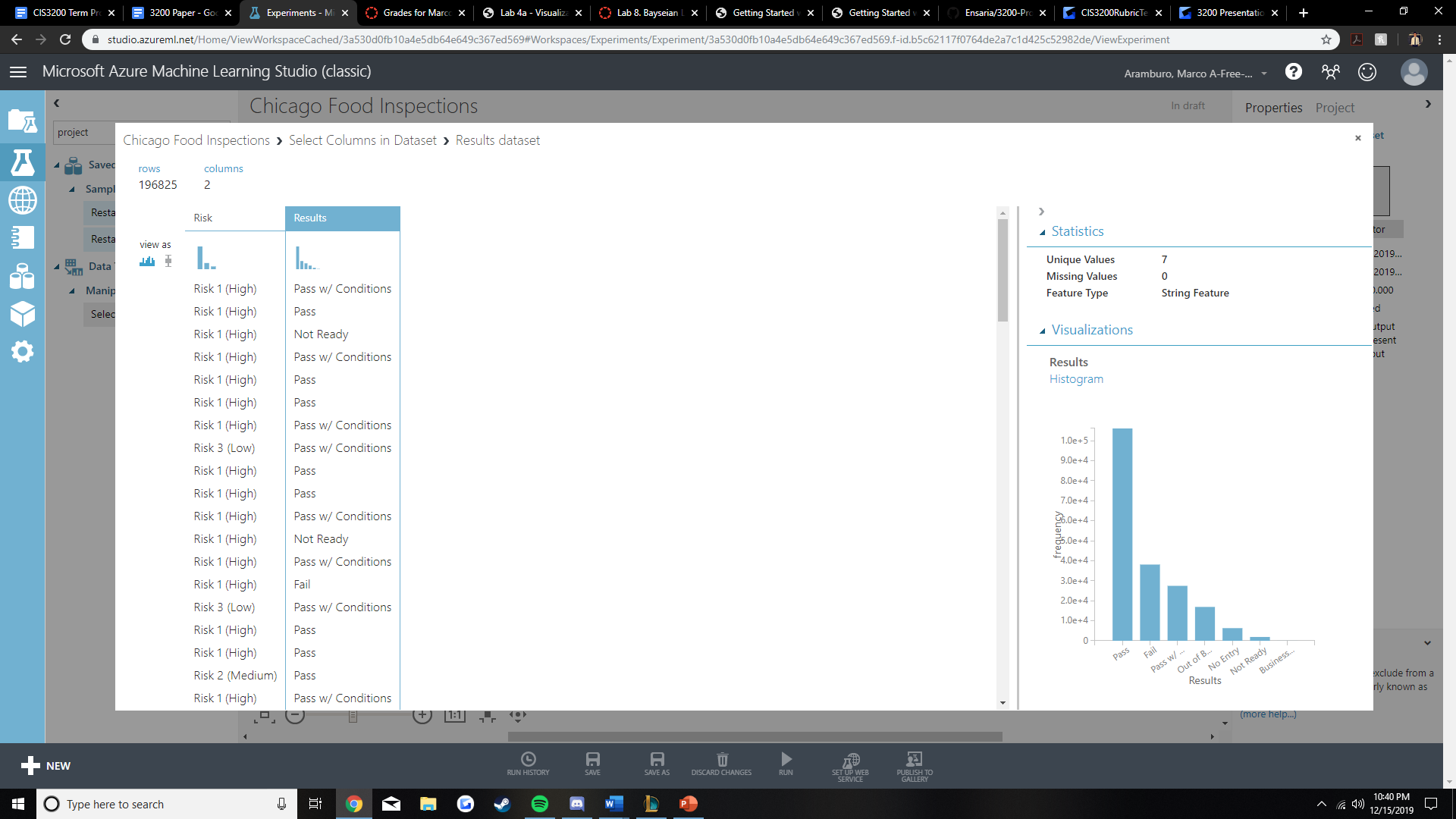
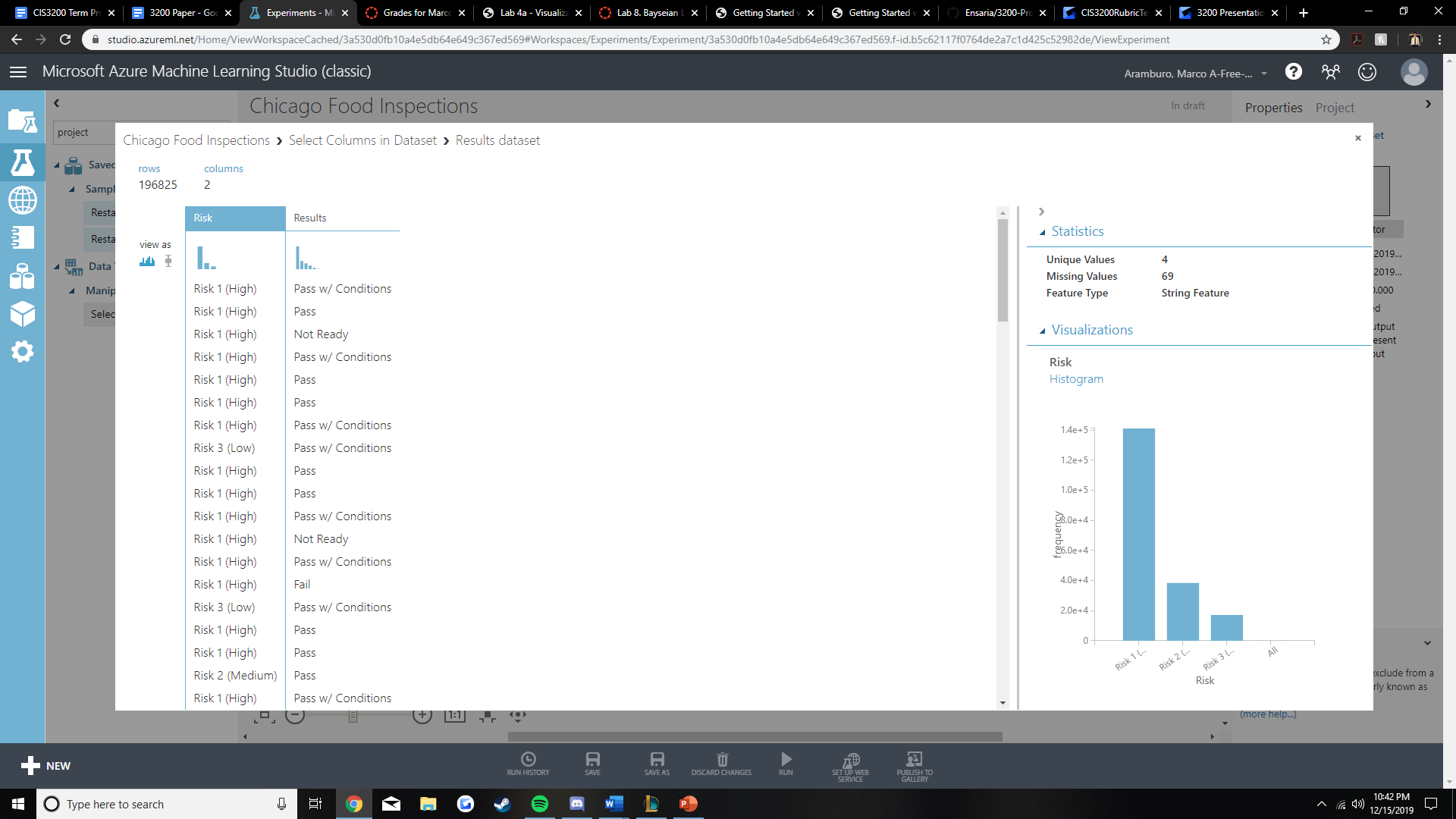
1. Under my datasets select and drag food inspections.cvs into the canvas
2. Then search for project columns also known as the select columns in dataset module in the search box.
3. Drag the select columns module onto the canvas and connect it to the data set. Connect both the data set to the column in dataset
4. Once connected search for R execute r script and connect it to the select columns in dataset module
5. Once connected locate Execute python-script module and drag it into the canvas ad connect it to the R execute module
6. Save all your work
7. Then search for Apply sql transformation module and drag it over into the canvas. Connecting it to Execute Python script
8. Locate a Split data module and drag it into the canvas.
9. Connect it to Apply SQL transformation module
10. Search for a Bayesian linear regression module and also a train model module
11. Connect the split module to the train module. Followed by connecting the Bayesian linear regression module to the train model.
12. Save your work
13. Search for a score model module and drag it into the canvas and connect it to the train model
14. Lastly search for the evaluate model module and connect it to the score module
15. And save & run

Step 3: Visualization

**Now to visualize the model. Here is one of the charts that you will be able to get once you run your model.**

**If you click on the columns, the graph will pop up in visuals to the right. If you hover over the bars it will show you the percentages of results.**

**For example, in the food that is being handled, there is a 72% risk. And for the results column only 54% of food distributors pass their inspections.**



References

1. <https://www.kaggle.com/chicago/chicago-food-inspections>

2. <https://github.com/Ensaria/3200-Project>

3. <https://studio.azureml.net>