

# Data Science and Machine Learning Essentials



Getting Started with MS Azure and Bayesian Linear Regression  
With Microsoft Azure Machine Learning Studio

Tutorial Project by Mario Arguello, Brianne Phillips, Michael Brumwell, and Christian Rodriguez  
CIS 3200 - Dr. Jongwook Woo

## Overview

In this tutorial you will know how to create an experiment, load the data set provided, and use regressions to calculate evaluations levels with Microsoft Azure Machine Learning Studio software. Our project consists of patient flu vaccinations from hospitals in every county of California. This tutorial is focused for students who are in beginner's level.

## What You'll Need

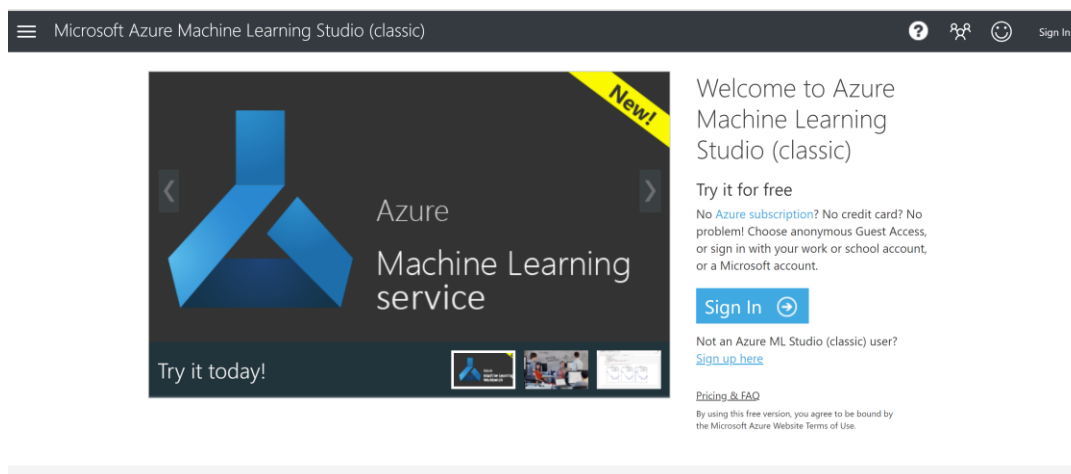
To complete this tutorial, you will need:


- Microsoft Azure Machine Learning Studio with a created account by clicking the website <https://studio.azureml.net/>
- You need to download the HCP Flu Vaccination data set by clicking this link <https://data.ca.gov/dataset/health-care-personnel-influenza-vaccination>
- Web browser with internet connection.

Make sure prior to using this tutorial please create an account for MS Azure ML Studio. It is free for one year. You can check out types of projects used for this powerful software used by real world Data Scientists and Machine Learning Engineers. Always be enthusiastic as these are fundamentals of entering this amazing field of data science.

## Create an Azure ML Account

1. When you log in to your Azure account click in the data set link provided above and download the CSV file in second data set called HCP Influenza Vaccination by Hospital and County.





**California Department of Public Health**





Social

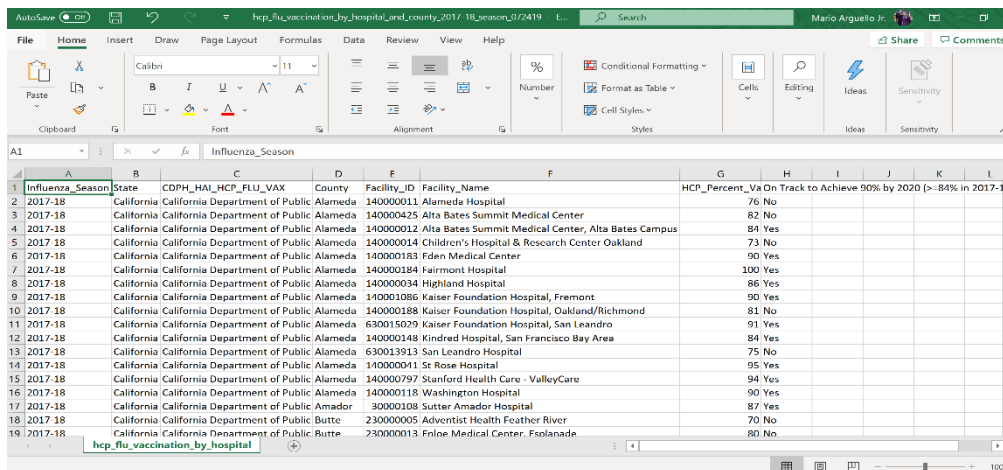
- Twitter
- Facebook
- LinkedIn

California acute care hospitals are required to offer free influenza vaccine to HCP. Hospital HCP must receive an annual vaccine or sign a declination form. Hospitals collect vaccination data for all HCP physically working in the hospital for at least one day during influenza season, regardless of clinical responsibility or patient contact. Hospitals report HCP vaccination rates to the California Department of Public Health (CDPH) and CDPH publishes the hospital results annually. CDPH reports data separately for hospital employees, licensed independent practitioners such as physicians, other contract staff, and trainees and volunteers (Health and Safety Code section 1288.7-1288.8).

Health and Safety Code section 1288.7(a) requires California acute care hospitals to offer influenza vaccine free of charge to all healthcare providers (HCP) or sign a declination form if a HCP chooses not to be vaccinated. Hospitals must report HCP influenza vaccination data to the California Department of Public Health (CDPH), including the percentage of HCP vaccinated. CDPH is required to make this information public on an annual basis [Health and Safety Code section 1288.8 (b)].

**Data and Resources**

-  **HCP Influenza Vaccination by County 2017-2018**  
This table shows the data for health care personnel (HCP) influenza... [Explore](#)
-  **Data Dictionary-HCP Influenza Vaccination by ...**  
Data Dictionary for HCP Influenza Vaccination by County 2017-2018 data file [Explore](#)
-  **HCP Influenza Vaccination by Hosp & County ...**  
This table shows the data for health care personnel (HCP) influenza... [Explore](#)
-  **Data Dictionary-HCP Influenza Vaccination by ...**  
Data Dictionary for HCP Influenza Vaccination by Hospital and County... [Explore](#)

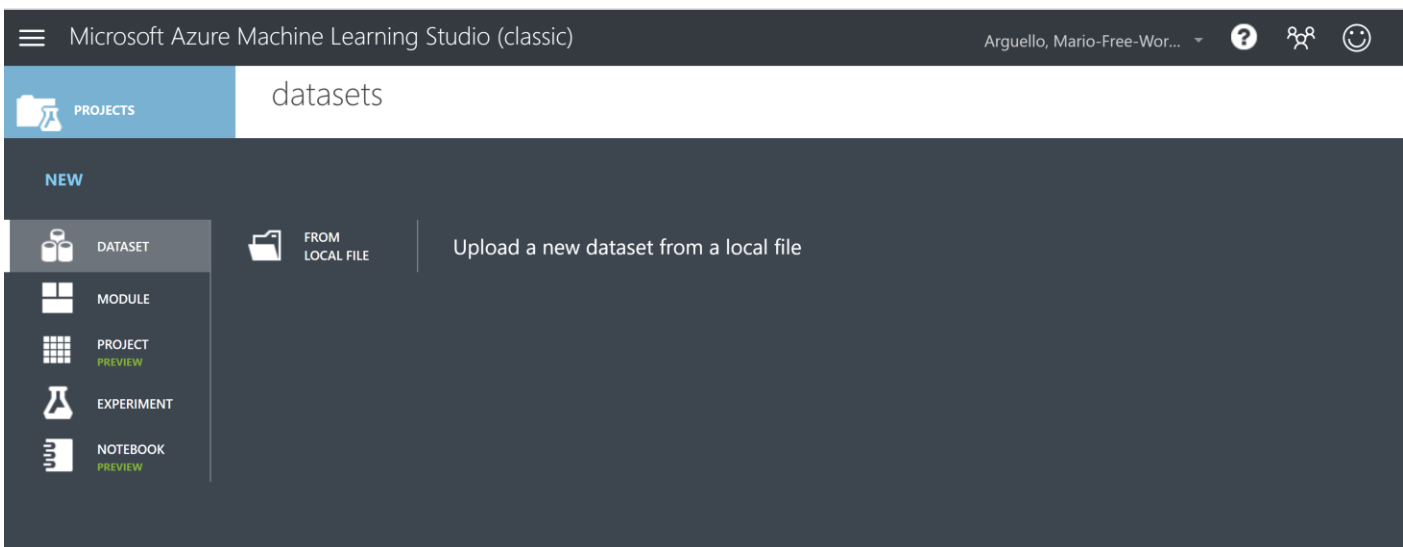


Influenza_Season	State	CDPH_HAI_HCP_FLU_VAX	County	Facility_ID	Facility_Name	HCP_Percent_Va On Track to Achieve 90% by 2020 (>=84% in 2017-18)
2017-18	California	California Department of Public Health	Alameda	140000011	Alameda Hospital	76 No
2017-18	California	California Department of Public Health	Alameda	140000425	Alta Bates Summit Medical Center	82 No
2017-18	California	California Department of Public Health	Alameda	140000012	Alta Bates Summit Medical Center, Alta Bates Campus	84 Yes
2017-18	California	California Department of Public Health	Alameda	140000014	Children's Hospital & Research Center Oakland	73 No
2017-18	California	California Department of Public Health	Alameda	140000183	Fden Medical Center	90 Yes
2017-18	California	California Department of Public Health	Alameda	140000194	Fairmont Hospital	100 Yes
2017-18	California	California Department of Public Health	Alameda	140000034	Highland Hospital	86 Yes
2017-18	California	California Department of Public Health	Alameda	140001086	Kaiser Foundation Hospital, Fremont	90 Yes
2017-18	California	California Department of Public Health	Alameda	140000188	Kaiser Foundation Hospital, Oakland/Richmond	81 No
2017-18	California	California Department of Public Health	Alameda	630015029	Kaiser Foundation Hospital, San Leandro	91 Yes
2017-18	California	California Department of Public Health	Alameda	140000148	Kindred Hospital, San Francisco Bay Area	84 Yes
2017-18	California	California Department of Public Health	Alameda	630013913	San Leandro Hospital	75 No
2017-18	California	California Department of Public Health	Alameda	140000041	St Rose Hospital	95 Yes
2017-18	California	California Department of Public Health	Alameda	140000797	Stanford Health Care - ValleyCare	94 Yes
2017-18	California	California Department of Public Health	Alameda	140000118	Washington Hospital	90 Yes
2017-18	California	California Department of Public Health	Amador	30000108	Sutter Amador Hospital	87 Yes
2017-18	California	California Department of Public Health	Butte	230000005	Adventist Health Feather River	70 No
2017-18	California	California Department of Public Health	Butte	230000013	Enloe Medical Center, Eureka	80 No

2. After you downloaded it the data set should look like this.

## Create an Experiment and Add Modules

- Let's start by clicking on data sets and click new data set located on the bottom left. Then click form local file to retrieve the data set HCP Flu Vaccination from which you saved it in.



Microsoft Azure Machine Learning Studio (classic)

Arguello, Mario-Free-Wor...

**PROJECTS** datasets

**NEW**

**DATASET** **FROM LOCAL FILE** Upload a new dataset from a local file

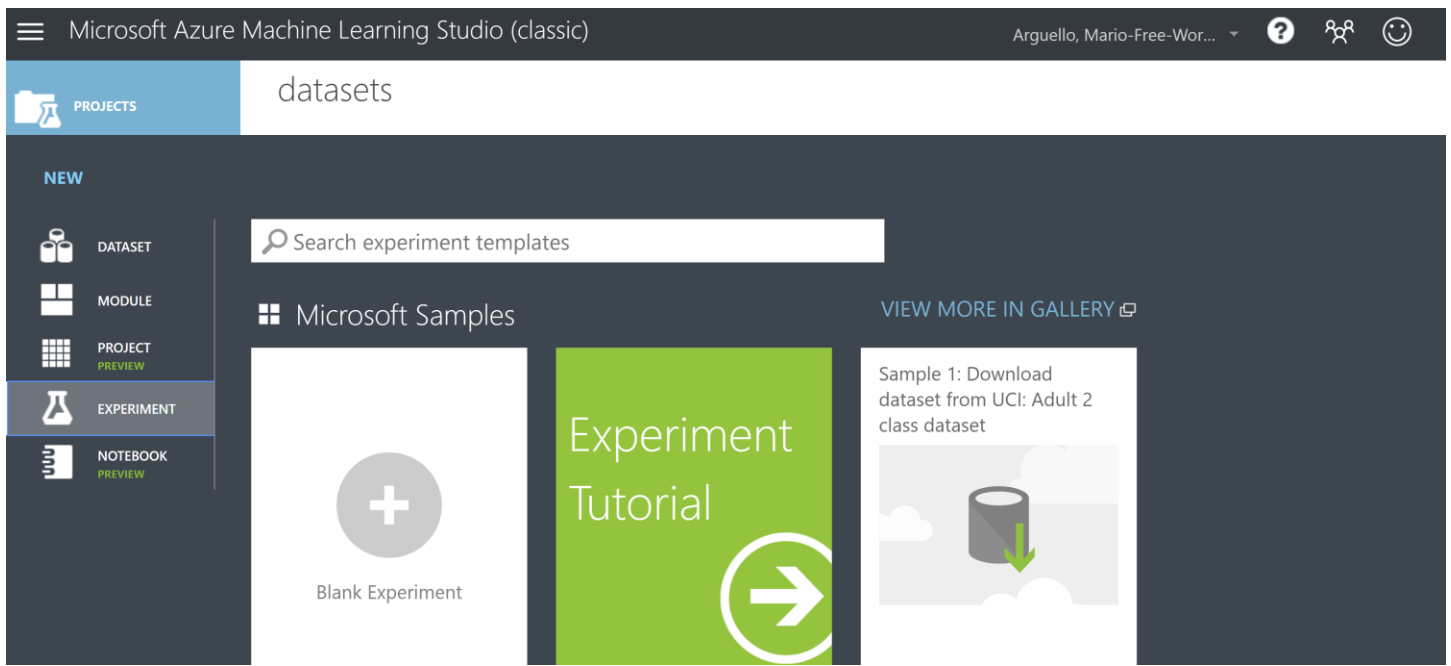
**MODULE**

**PROJECT PREVIEW**

**EXPERIMENT**

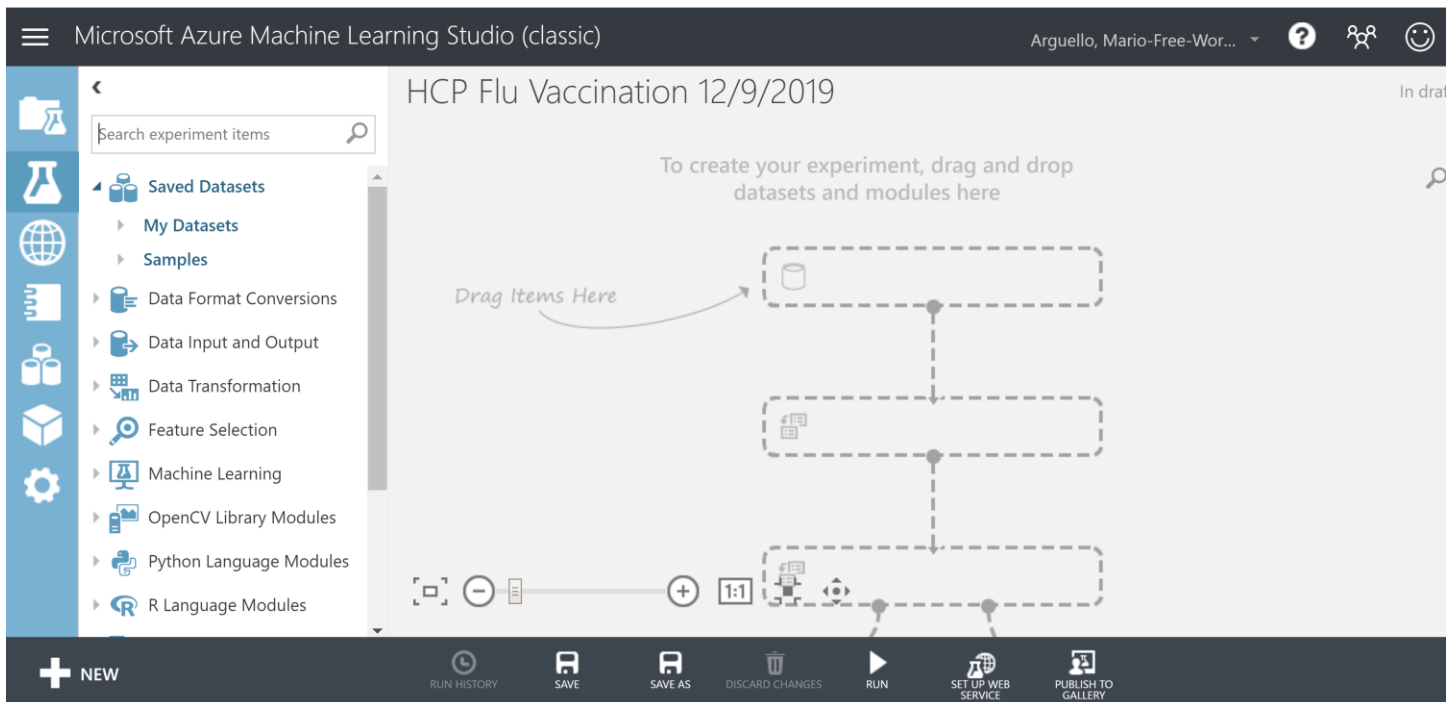
**NOTEBOOK PREVIEW**

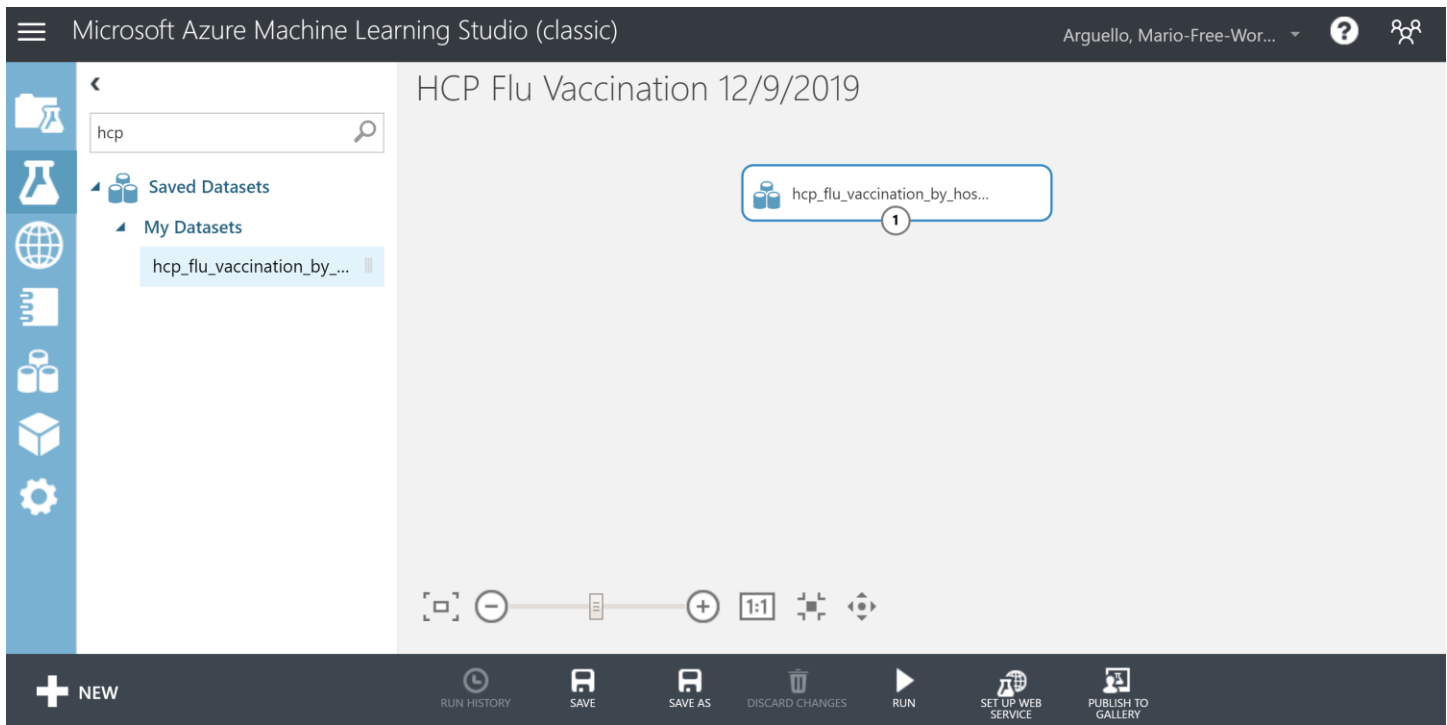
- After the data set is uploaded let's create a new experiment by clicking on the experiment → blank experiment in the same column where you clicked dataset.



## Uploading a Data File to Azure ML

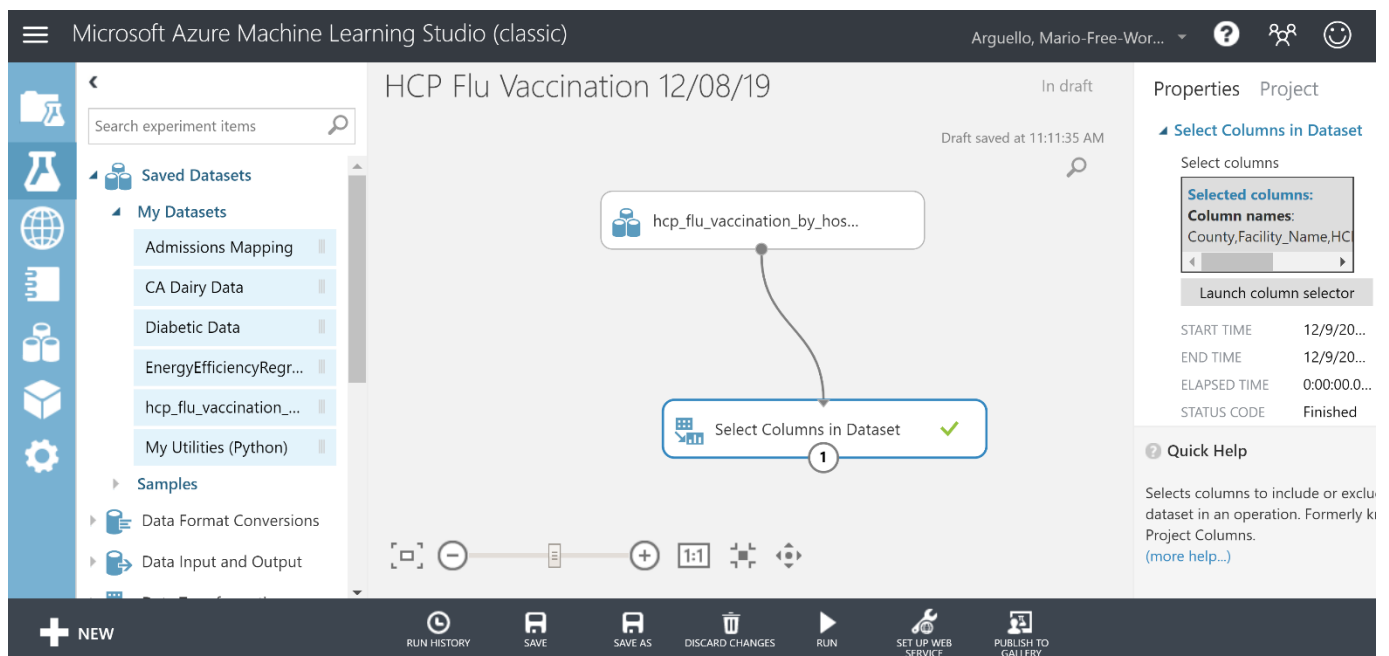
- When you open the new blank experiment name the experiment **HCP Flu Vaccination**. Now search **HCP Flu Vaccination** on the top left search bar and drag the data set module to the center. Your work should resemble the images below.





## Creating an Azure ML Experiment

- Now that you are done with these steps search **Select Columns in Data Set** and drag it to the center such as the first module you dragged for the data set. Launch the column selector → click by name → select County, Facility Name, and HCP\_Percent\_Vaccinated. Connect the HCP flu data to the **Select Columns Data Set**. Make sure to save and run the experiment. The modules should resemble the picture below.



- Now search for **Split Data** and drag them to the center. At the properties pane under split data adjust the following; splitting rows: Split Rows, fraction of rows: 0.5, random speed: 0. Under train model at the properties pane → launch column selector → click with rules → search HCP\_Percent\_Vaccinated. Click the check box on the properties pane.
- Connect from the **Select Columns Data Set** to the **Split Data** module.

9. Now search **Bayesian Linear Regression** and drag it to the center. At the properties pane on the right make sure that the regularization weight is 1 and checked box on allow unknown.
10. Always make sure to save. Do not run yet until the next step is completed.

Microsoft Azure Machine Learning Studio (classic)

HCP Flu Vaccination 12/08/19

In draft

Draft saved at 11:31:31 AM

Search experiment items

Saved Datasets

- My Datasets
  - Admissions Mapping
  - CA Dairy Data
  - Diabetic Data
  - EnergyEfficiencyRegr...
  - hcp\_flu\_vaccination\_...
  - My Utilities (Python)
- Samples
  - Data Format Conversions
  - Data Input and Output

Properties Project

Experiment Properties

START TIME	12/9/20.
END TIME	12/9/20.
STATUS CODE	InDraft
STATUS DETAILS	None

Prior Run

Summary

Enter a few sentences describing your experiment (up to 140 characters).

Quick Help

NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

11. Search **Train Model** module and drag it. Under train model at the properties pane → launch column selector → click with rules → search HCP\_Percent\_Vaccinated. Then connect the **Bayesian Linear Regression** module to the left dot of the **Train Model**. Now connect **Split Data** to the **Train Model**. Save and Run the experiment.

Microsoft Azure Machine Learning Studio (classic)

HCP Flu Vaccination 12/08/19

Draft saved at 1

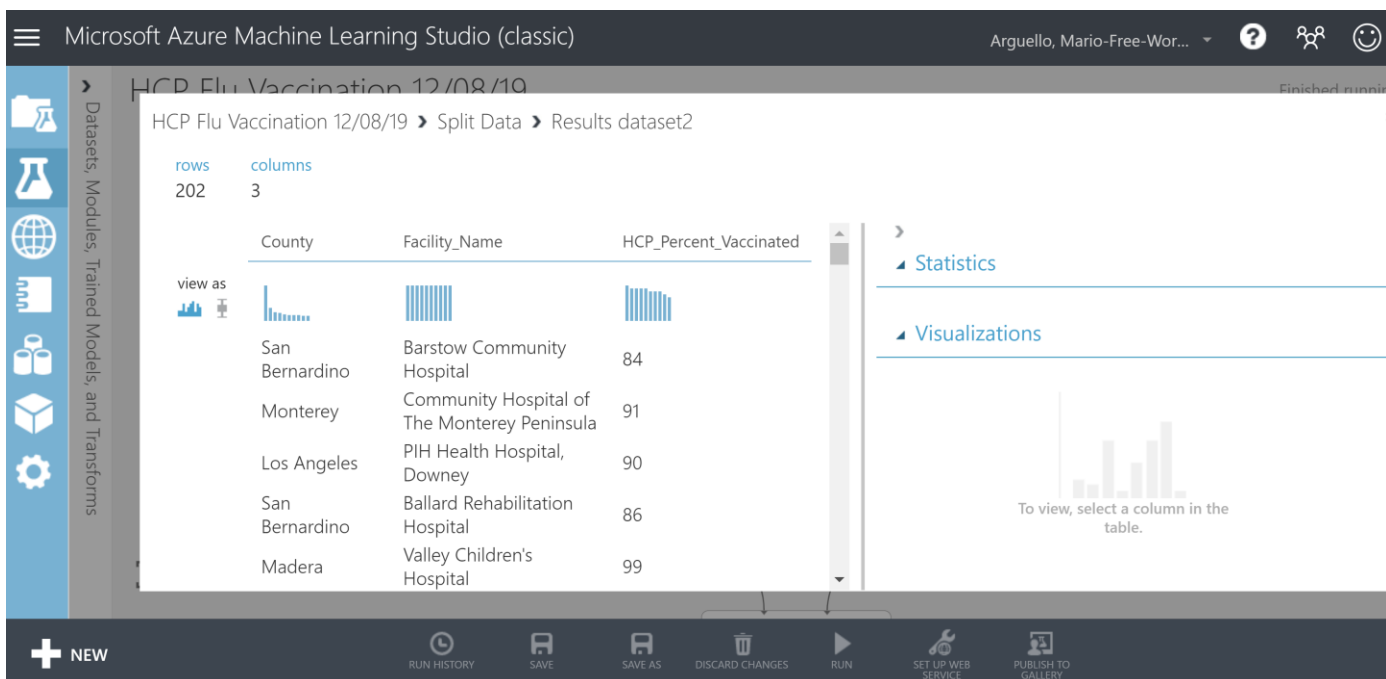
Datasets, Modules, Trained Models, and Transforms

Train Model

NEW RUN HISTORY SAVE SAVE AS DISCARD CHANGES RUN SET UP WEB SERVICE PUBLISH TO GALLERY

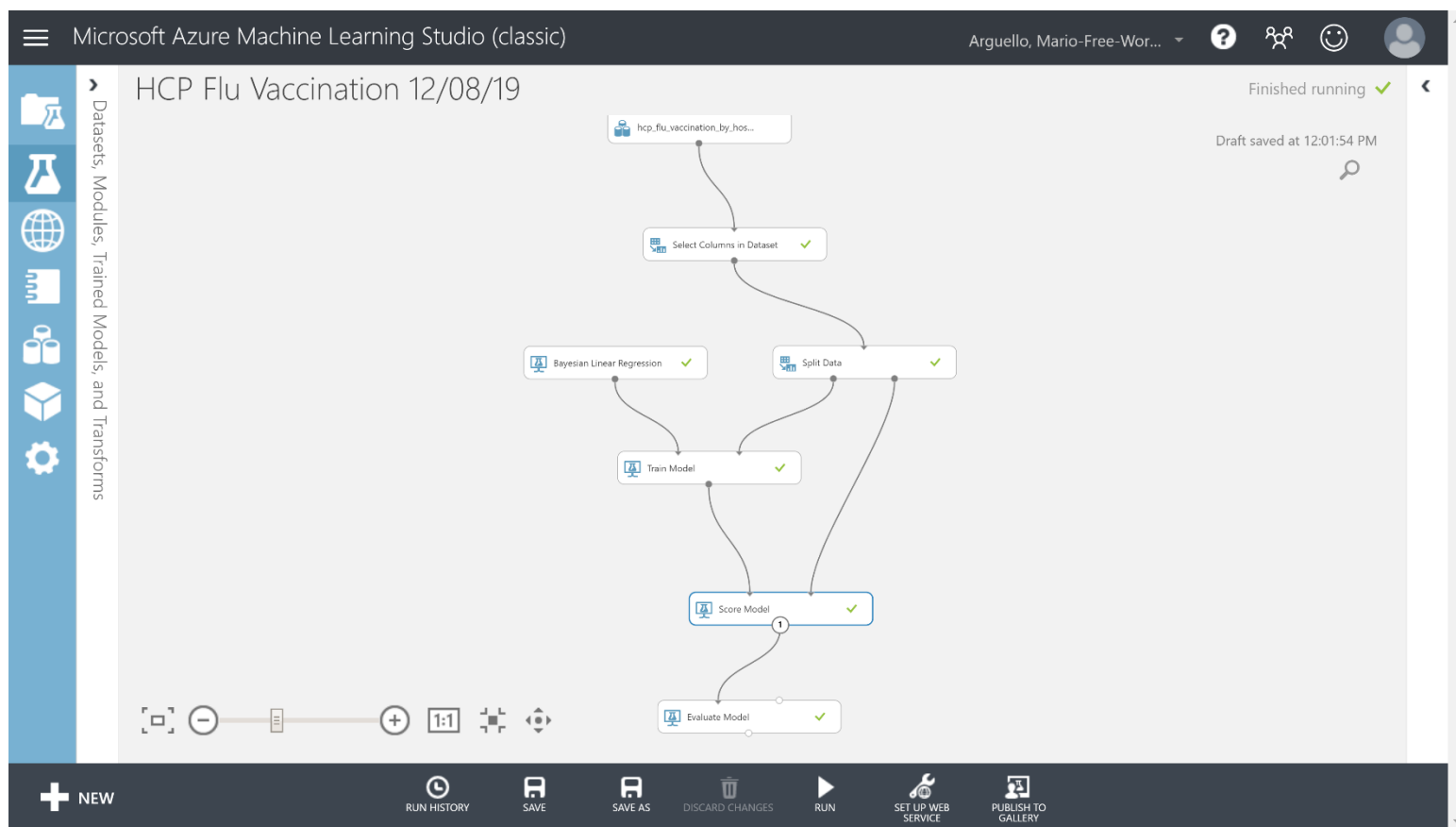
Examine the Data

12. Now let's right click **Split Data** → click on dataset 1 → click visualize. Do this with the data set 2. You can see some cool insights of the percentage of the patients who got vaccinated in the flu season of 2017-2018. These insights shows the county names, the names of the hospitals where the patients where vaccinated. Also the percentage. You are almost done with this tutorial. Refer the images below for accuracy.

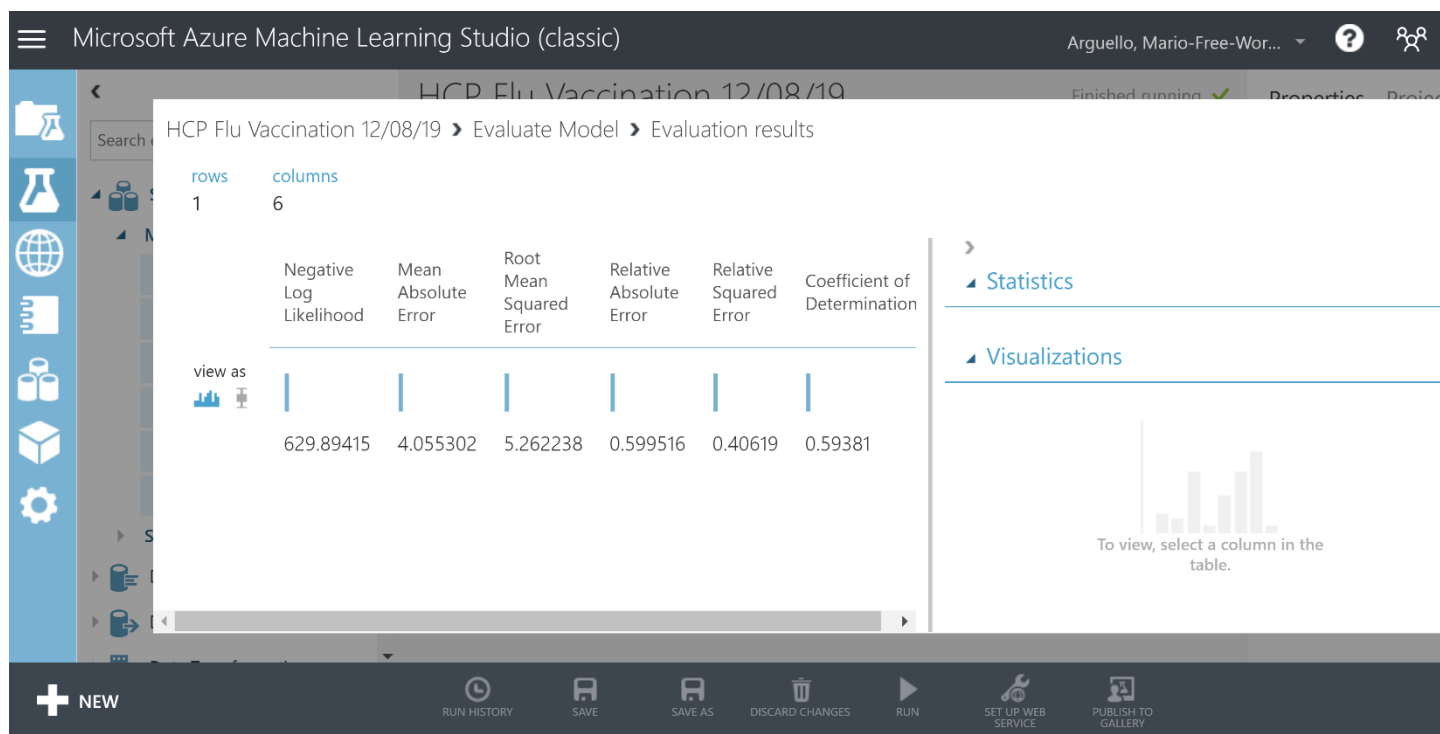


13. Now search **Score Model** and drag it to the center. Then search **Evaluate Model** and drag it to the center.
14. Now connect the **Train Model** module to the **Score Model**. Then connect from the right dot of **Split Data** to the **Score Model**.

15. Lastly, connect the **Score Model** module to the **Evaluate Model** module. The experiment should resemble the image below.



16. Save and run the experiment. You should see check marks every time you save and run.
17. Right click the **Evaluate Model** module → scroll to evaluation results → click visualize. you see the evaluation results of the linear regression of the percentages of patients who were vaccinated in the



hospitals of every county in California. This shows the margins of the statistics of this data set. You can see that there are root mean squared errors and relative absolute error. They have the numbers in decimal.

18. This concludes the tutorial of this project on Microsoft Azure Machine Learning Studio.

## Summary

In this tutorial you have learned how to use Microsoft Azure at a beginner's level. You also learned on how to upload the data set provided to this software tool. You also learned how to modify the properties pane to filter our this data set. The purpose of this tutorial project is to learn Bayesian linear regression models and using score and train models to obtain insights of evaluation results when the model is running.

## References

<https://data.ca.gov/dataset/health-care-personnel-influenza-vaccination>

California Department of Public Health Influenza Vaccination Data Set

<https://data.ca.gov/dataset>

California Department of Public Health open data website

<https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/bayesian-linear-regression>

Bayesian Linear Regression Concept