IRIS SPECIES

"Classify iris plants into three different species"

Jenalyn Christian D. Simbahan

Motis 8

Table of Contents

Iris Species

The Iris flower data set or Fisher's Iris data set is a multivariate data set introduced by Sir Ronald Aylmer Fisher (1936) as an example of discriminant analysis. It is sometimes called Anderson's Iris data set because Edgar Anderson collected the data to quantify the morphologic variation of Iris flowers of three related species. The dataset consists of 50 samples from each of three species of Iris flowers (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample, they are the length and the width of sepal and petal, in centimeters. Based on the combination of the four features, Fisher developed a linear discriminant model to distinguish the species from each other.

Requirements

To analyze the data set using the Daitaiku tool which is a Collaborative Data Science Platform and to classify the Iris plan into 3 species from measurements of their petals and sepal and to share the analysis about the data set.

Supporting data

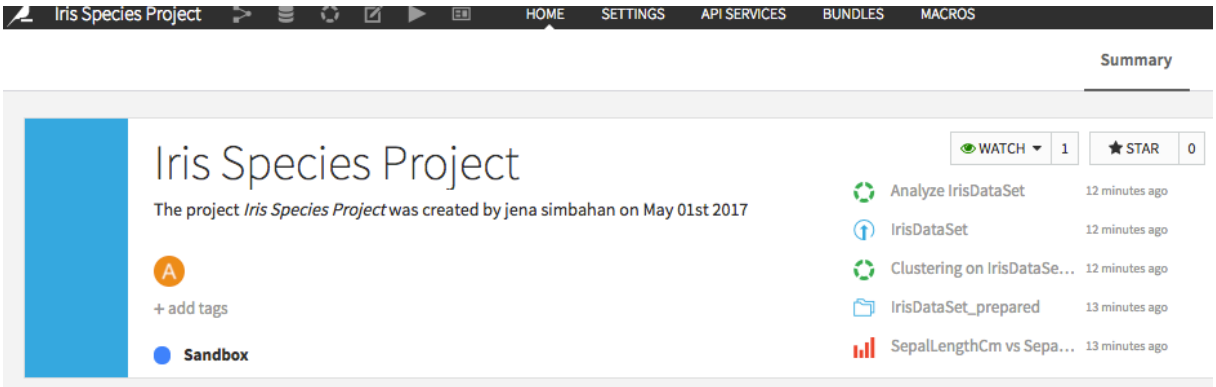
- The dataset: iris.csv

The columns in this dataset are composed of:

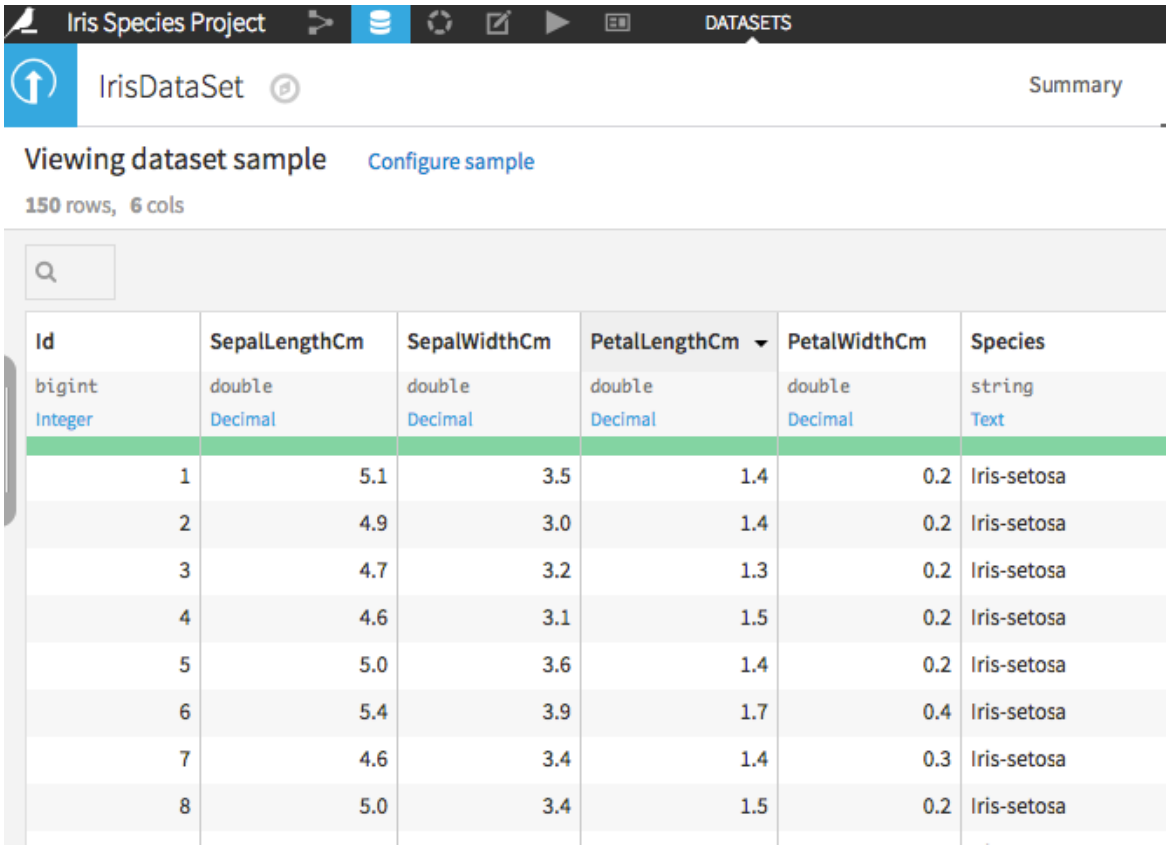
- Id
- SepalLengthCm
- SepalWidthCm
- PetalLengthCm
- PetalWidthCm
- Species

Detailed Process

- Created the project of **Iris Species Project**



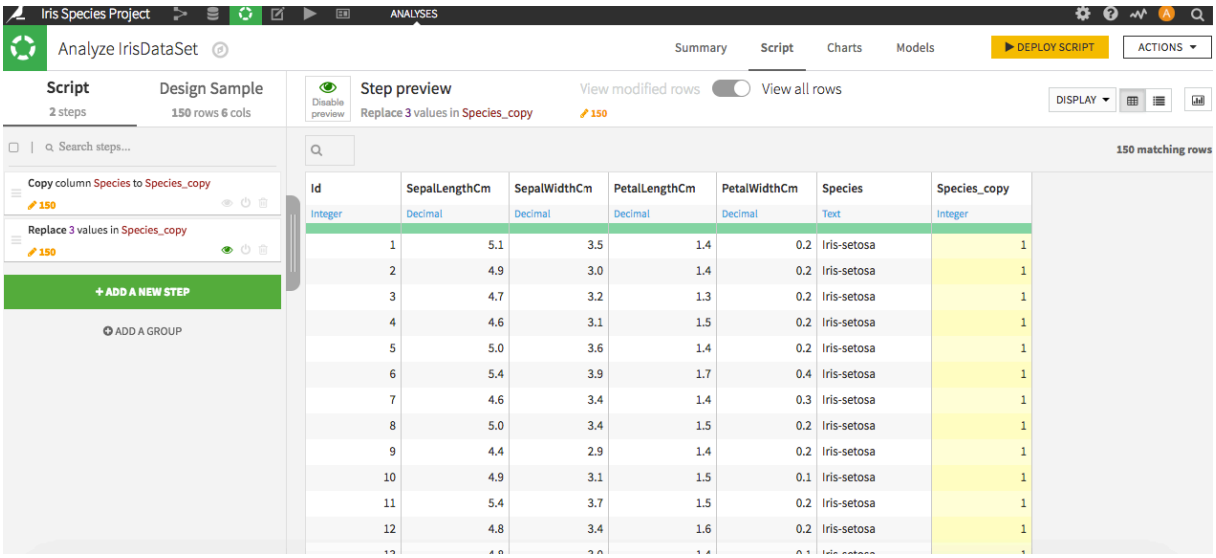
- Upload the **Iris dataset** (iris.csv). It provides the dimensions of the data set and it says it has 150 rows and 6 columns.



Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
bigint	double	double	double	double	string
Integer	Decimal	Decimal	Decimal	Decimal	Text
1	5.1	3.5	1.4	0.2	Iris-setosa
2	4.9	3.0	1.4	0.2	Iris-setosa
3	4.7	3.2	1.3	0.2	Iris-setosa
4	4.6	3.1	1.5	0.2	Iris-setosa
5	5.0	3.6	1.4	0.2	Iris-setosa
6	5.4	3.9	1.7	0.4	Iris-setosa
7	4.6	3.4	1.4	0.3	Iris-setosa
8	5.0	3.4	1.5	0.2	Iris-setosa

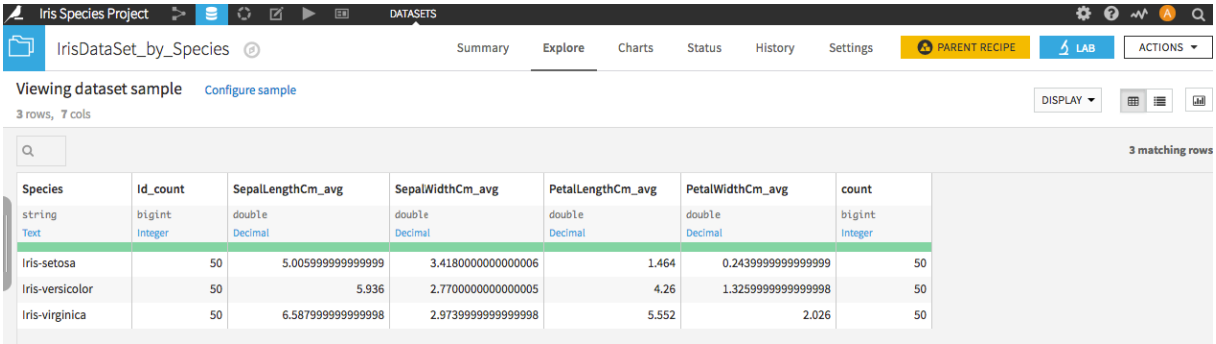
- While exploring the data set, the head command gives the first 6 rows of the data sets. It is observed that it has 6 columns with header names which are: SepalLength,SepalWidth,PetalLength,PetalWidth and Species. Species is one categorical type. The data set has data of 150 flowers. The data is composed of 50 setosa flowers, 50 versicolor flowers and 50 virginica flowers.

- Data preparation and data cleansing



Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species	Species_copy
Integer	Decimal	Decimal	Decimal	Decimal	Text	Integer
1	5.1	3.5	1.4	0.2	Iris-setosa	1
2	4.9	3.0	1.4	0.2	Iris-setosa	1
3	4.7	3.2	1.3	0.2	Iris-setosa	1
4	4.6	3.1	1.5	0.2	Iris-setosa	1
5	5.0	3.6	1.4	0.2	Iris-setosa	1
6	5.4	3.9	1.7	0.4	Iris-setosa	1
7	4.6	3.4	1.4	0.3	Iris-setosa	1
8	5.0	3.4	1.5	0.2	Iris-setosa	1
9	4.4	2.9	1.4	0.2	Iris-setosa	1
10	4.9	3.1	1.5	0.1	Iris-setosa	1
11	5.4	3.7	1.5	0.2	Iris-setosa	1
12	4.8	3.4	1.6	0.2	Iris-setosa	1

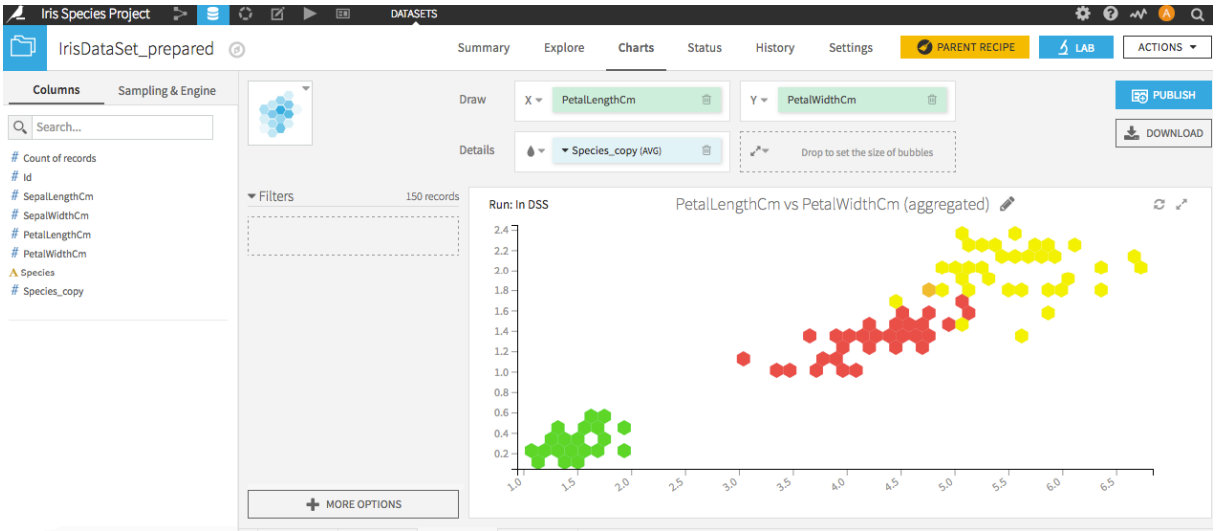
- Group per Iris species



Species	Id_count	SepalLengthCm_avg	SepalWidthCm_avg	PetalLengthCm_avg	PetalWidthCm_avg	count
string Text	bigint Integer	double Decimal	double Decimal	double Decimal	double Decimal	bigint Integer
Iris-setosa	50	5.005999999999999	3.418000000000006	1.464	0.2439999999999999	50
Iris-versicolor	50	5.936	2.770000000000005	4.26	1.3259999999999998	50
Iris-virginica	50	6.587999999999998	2.973999999999998	5.552	2.026	50

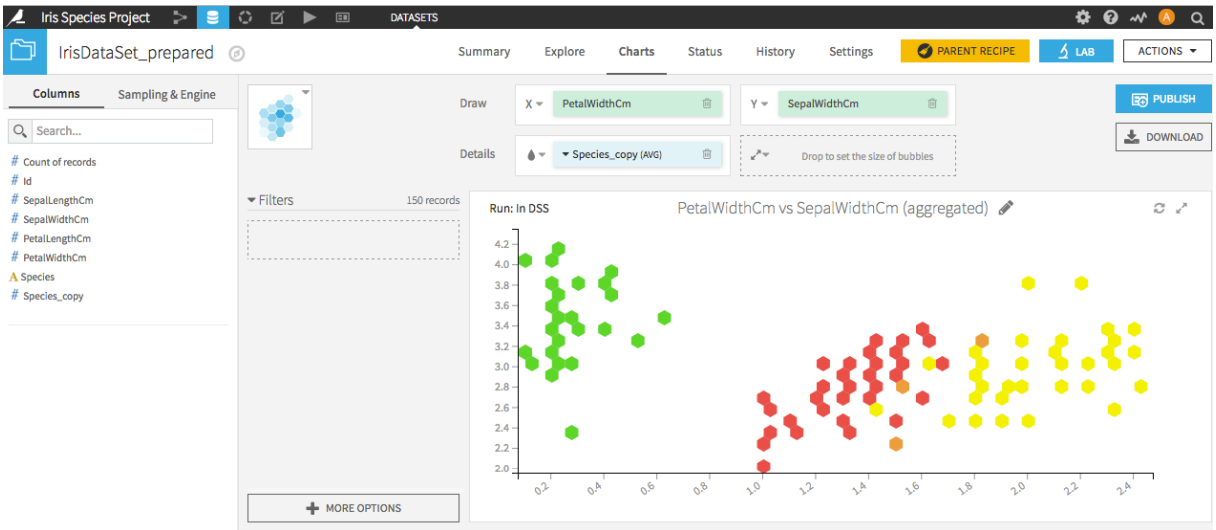
- Iris species by Sepal Length and Width

- Iris species by Petal Length and Width



- Iris species by Petal and Sepal Width

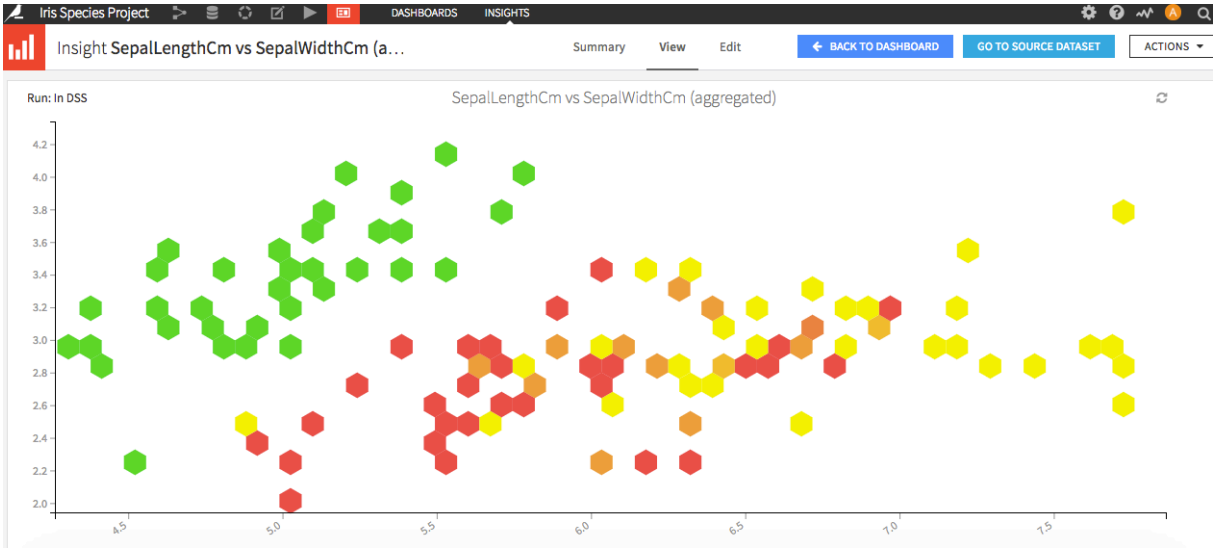
- So it is shows that petal width differs by species. It is visible that the petal width of virginica is more but this can be due to natural variability. From the plot also it is clear that the petal width of verginica is more than the versicolor and the thinnest of the petals are in setosa.



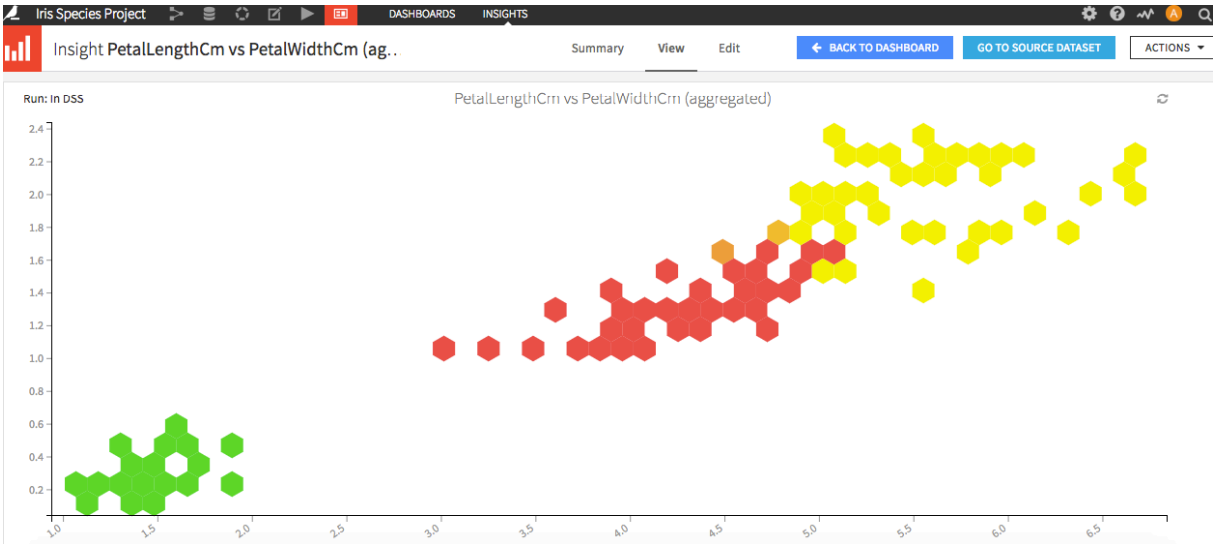
- Iris species by Petal and Sepal Length

- It is shown that the petal length increases with the sepal length, for whatever be the species. It is seen that petal length increases with the sepal length for versicolor and virginica species too. But for setosa the same is not much visible. Based on their values and plot only, there means seem clearly different.

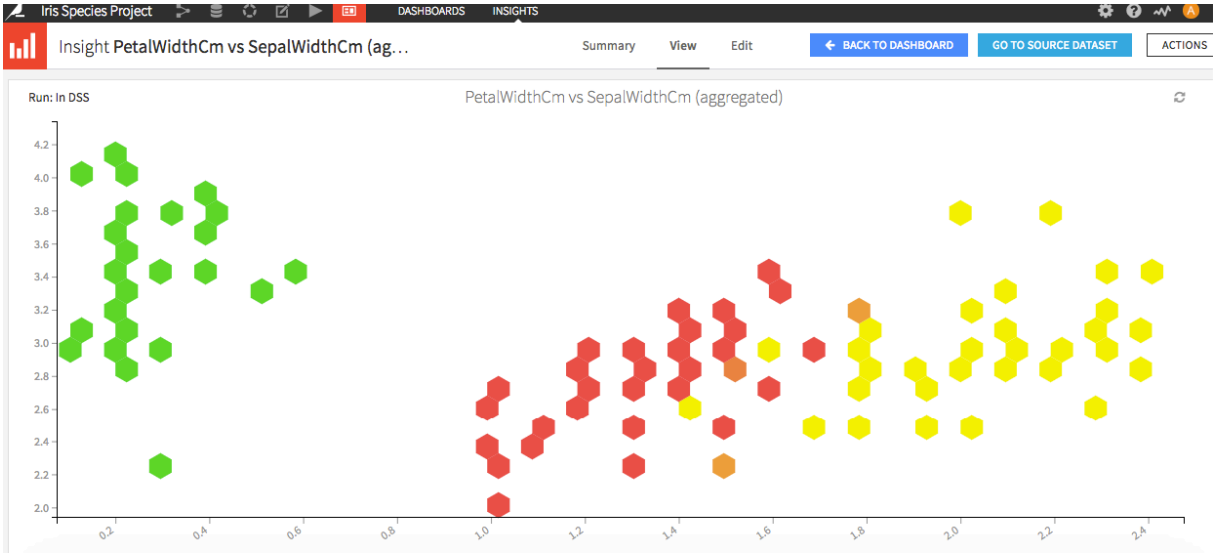
- Dashboard created for Iris species by Sepal Length and Width



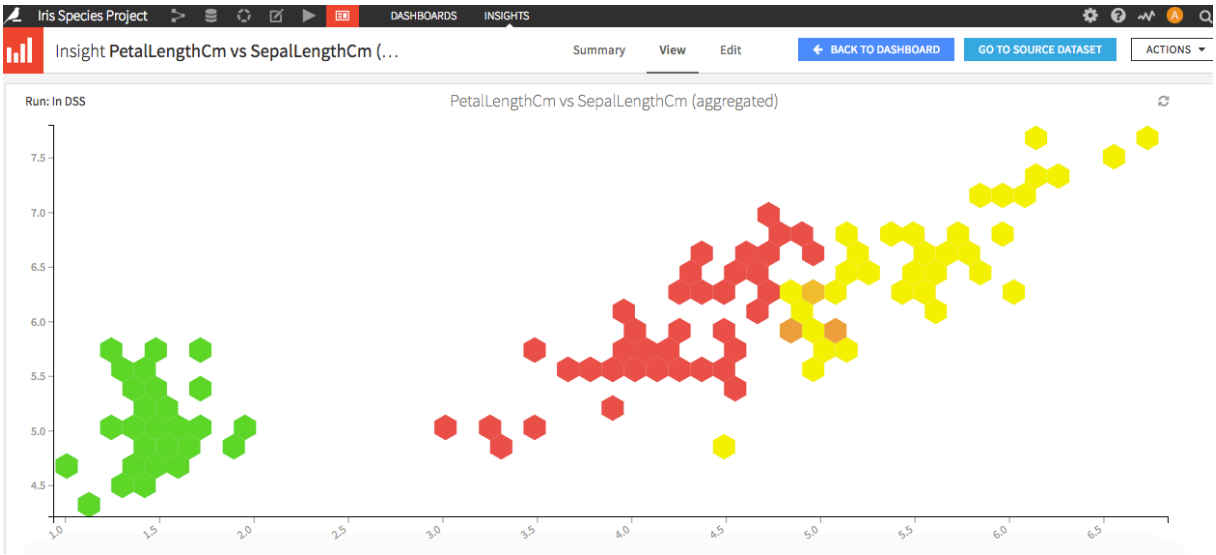
- Dashboard created for Iris species by Petal Length and Width.



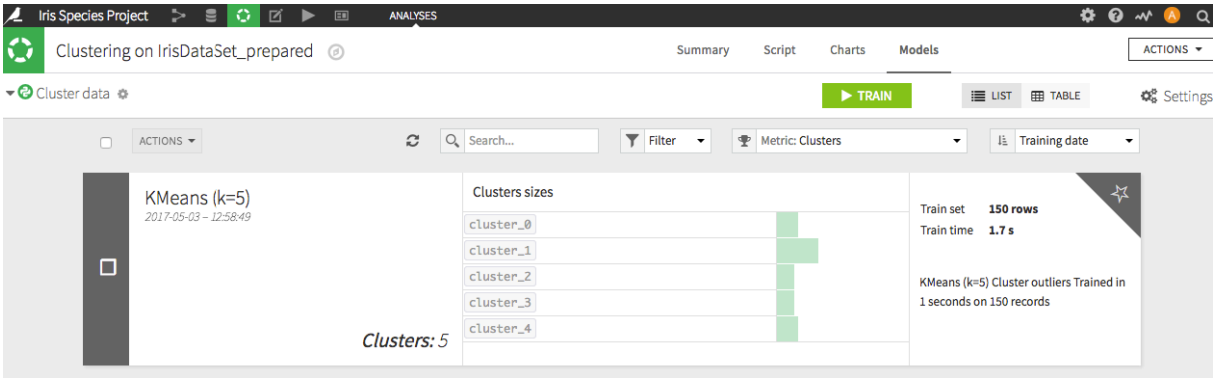
- Dashboard created for Iris species by Petal and Sepal Width.



- Dashboard created for Iris species by Petal and Sepal Length.



- Created my model and clustered the Iris species by using the unsupervised learning



- Please see below my workflow overview result