Project Development Phase
Model Performance Test

Date	10 November 2022
Team ID	PNT2022TMID42453
Project Name	Project -Estimate of crop yield production using data analytics
Maximum Marks	10 Marks

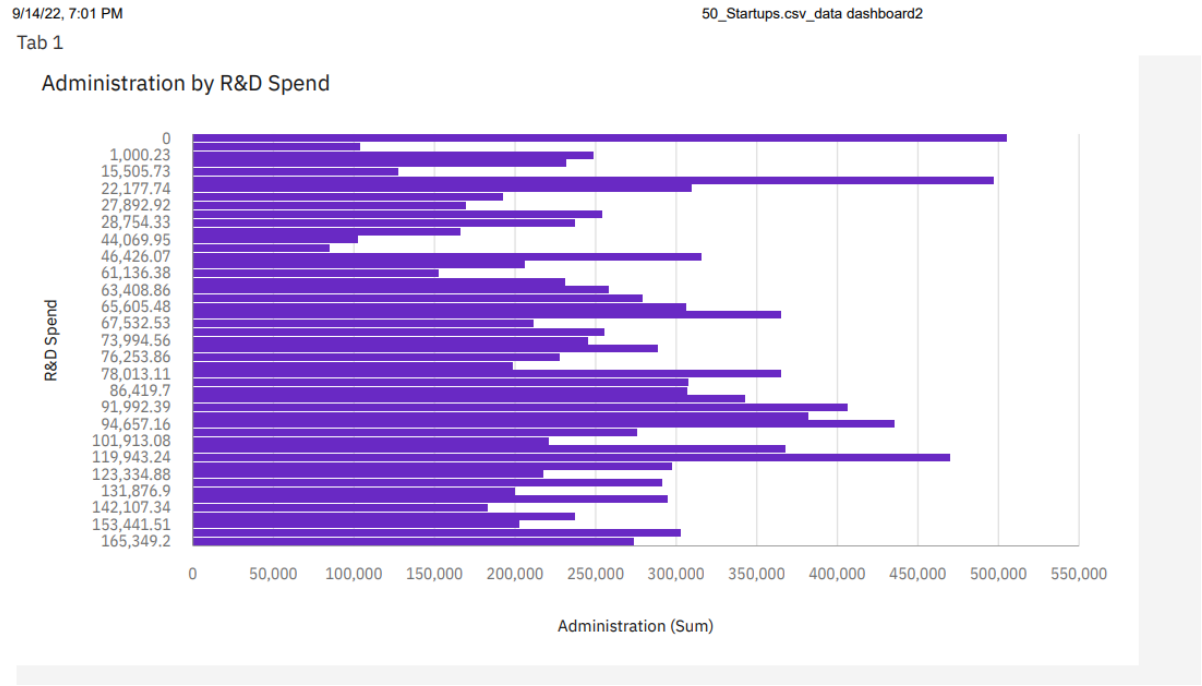
Model Performance Testing:

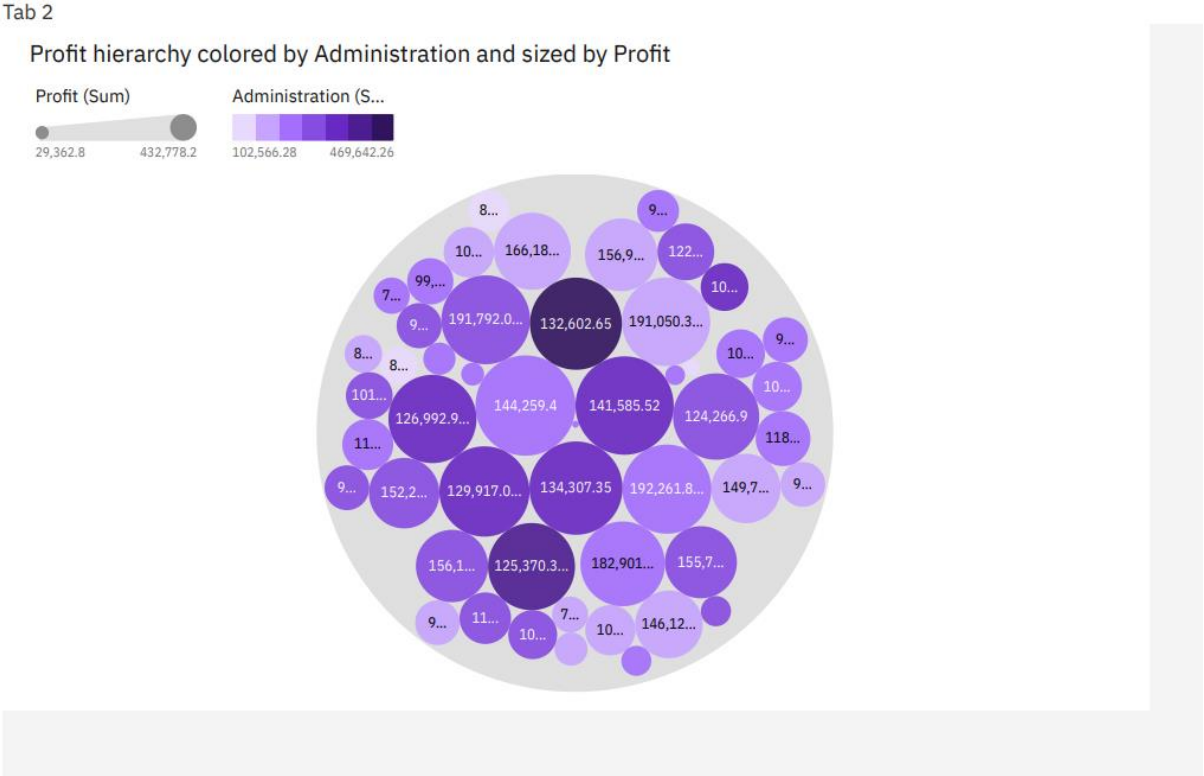
Project team shall fill the following information in model performance testing template.

S.No.	Parameter	Screenshot / Values
1.	Dashboard design	No of Visulizations =10
2.	Data Responsiveness	3
3.	Amount Data to Rendered (DB2 Metrics)	2
4.	Utilization of Data Filters	1
5.	Descriptive Reports	No of Visulizations=5

Dashboard design:

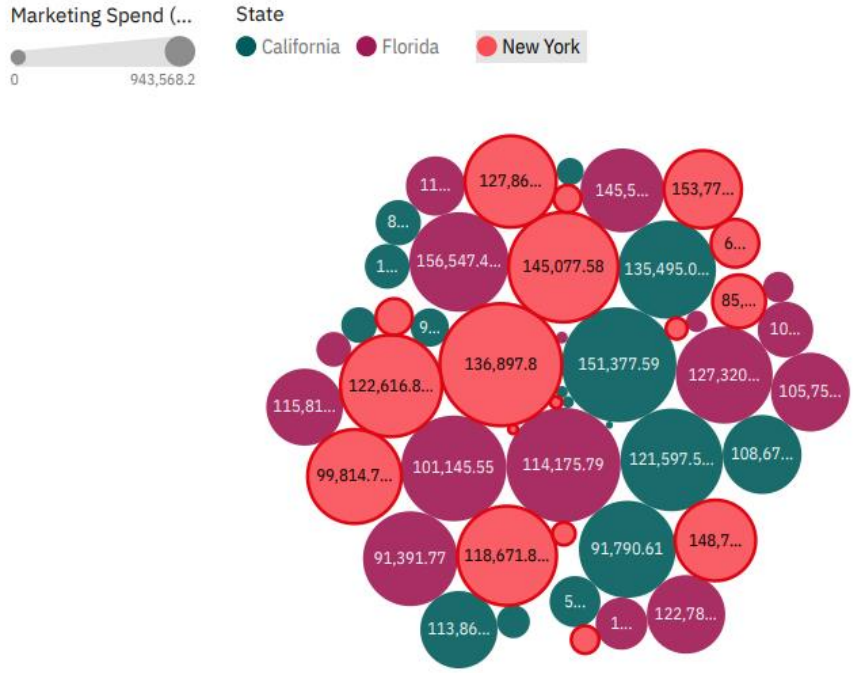
Number of Visulizations =10



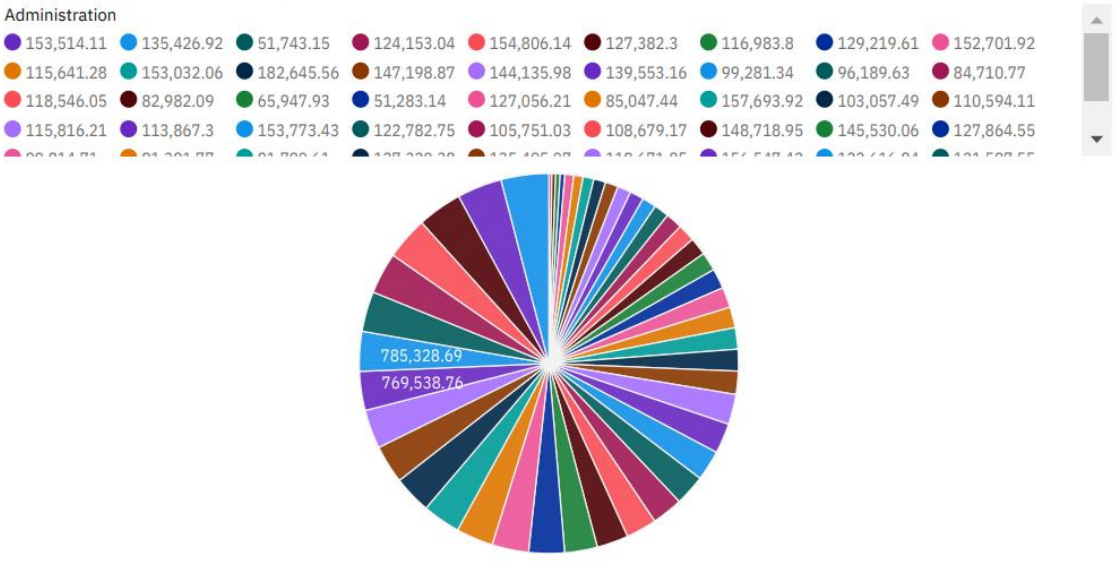


Tab 4

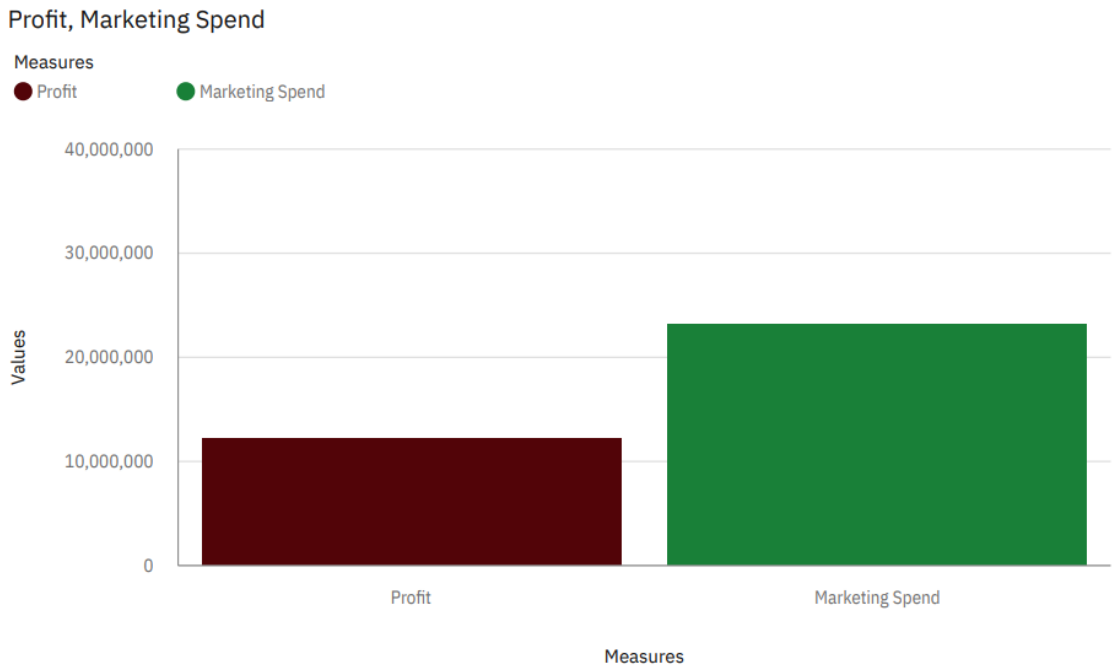
Administration colored by State sized by Marketing Spend



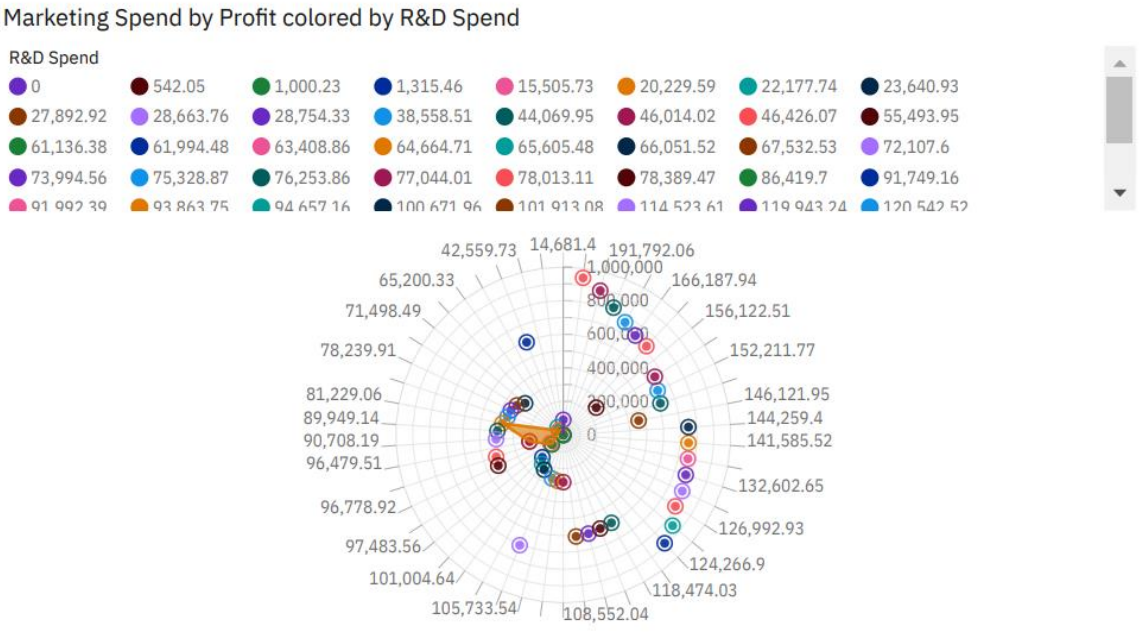
Marketing Spend by Administration



Tab 7


Profit, Marketing Spend
Measures
Profit
Marketing Spend
40,000,000
30,000,000
20,000,000
10,000,000
0
Values
Profit
Marketing Spend
Measures

Tab 6


Marketing Spend by Profit colored by R&D Spend
R&D Spend
0
542.05
1,000.23
1,315.46
15,505.73
20,229.59
22,177.74
23,640.93
27,892.92
28,663.76
28,754.33
38,558.51
44,069.95
46,014.02
46,426.07
55,493.95
61,136.38
61,994.48
63,408.86
64,664.71
65,605.48
66,051.52
67,532.53
72,107.6
73,994.56
75,328.87
76,253.86
77,044.01
78,013.11
78,389.47
86,419.7
91,749.16
42,559.73
14,681.4
191,792.06
1,000,000
65,200.33
166,187.94
71,498.49
156,122.51
78,239.91
152,211.77
400,000
81,229.06
146,121.95
89,949.14
144,259.4
90,708.19
141,585.52
96,479.51
132,602.65
96,778.92
126,992.93
97,483.56
124,266.9
101,004.64
118,474.03
105,733.54
108,552.04

Tab 9

R&D Spend

R&D Spend

29,592.7 103,829

Nodes

All

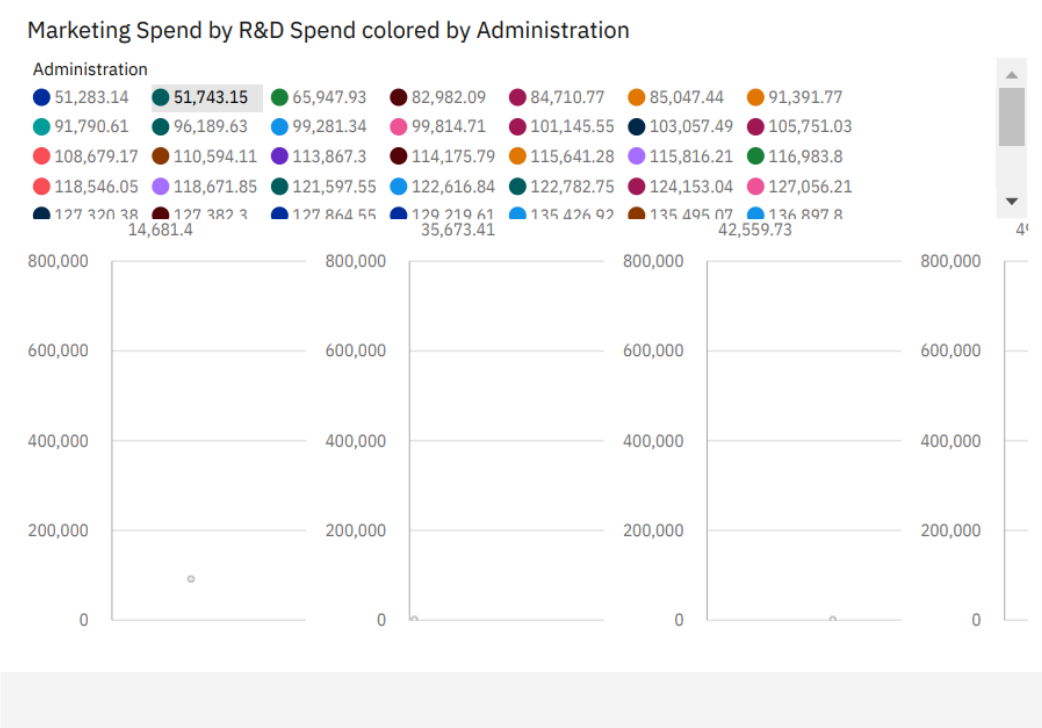
Tab 8

R&D Spend

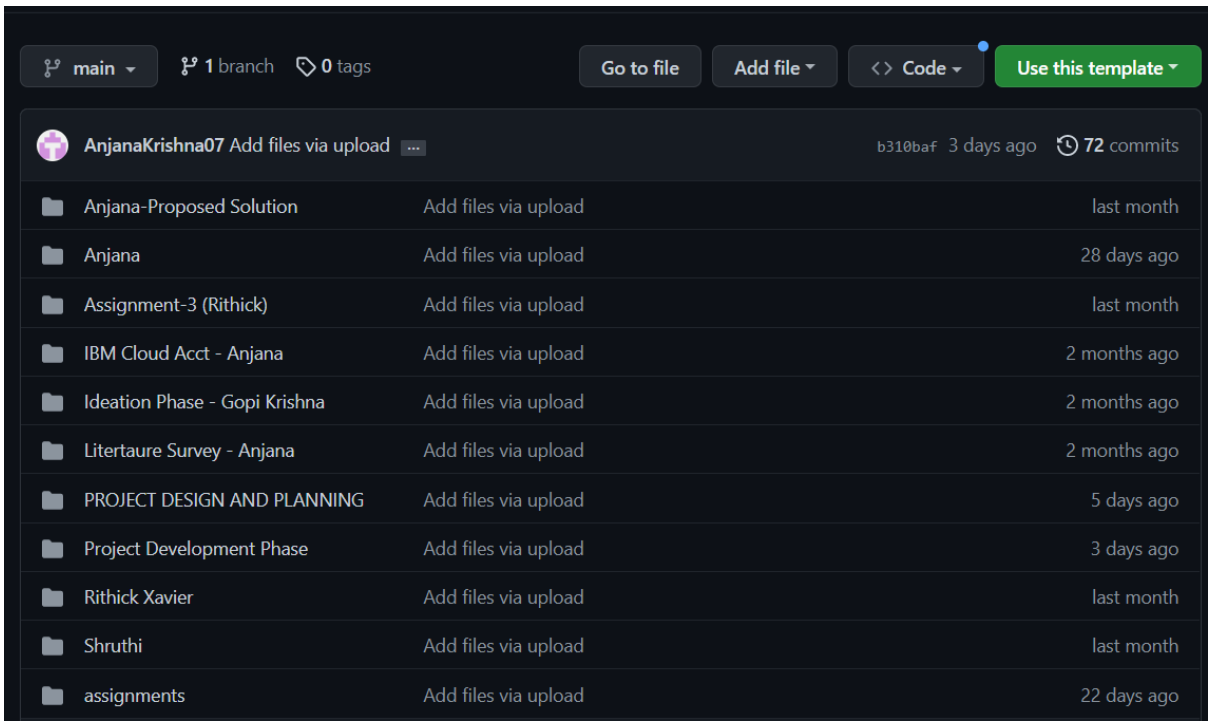
Column values

Increase Decrease Sum

Tab 10

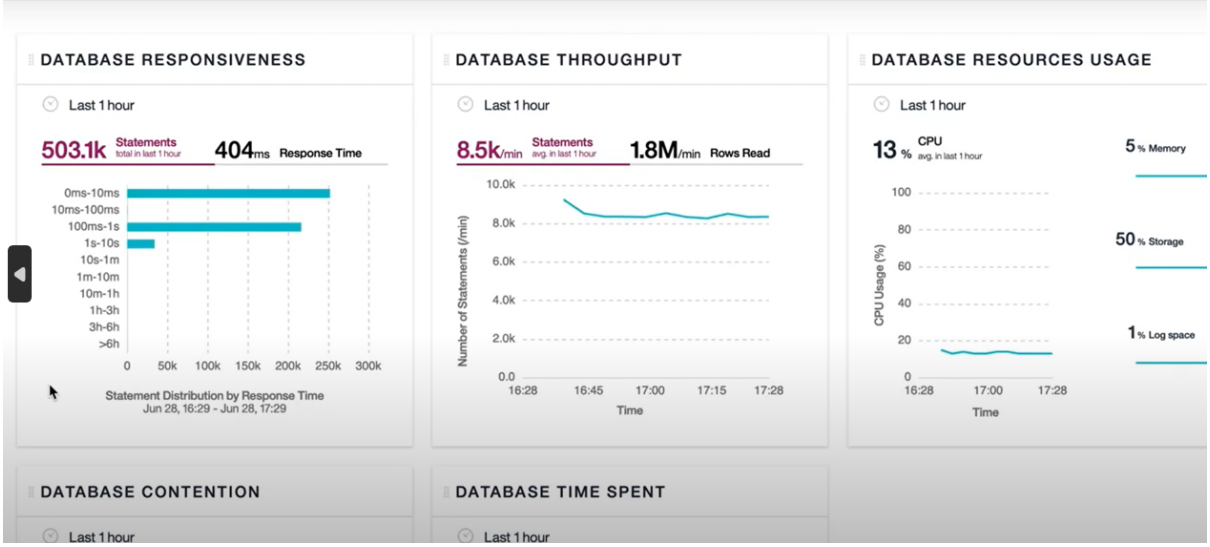

Marketing Spend by R&D Spend colored by Administration
Administration
51,283.14
51,743.15
65,947.93
82,982.09
84,710.77
85,047.44
91,391.77
91,790.61
96,189.63
99,281.34
99,814.71
101,145.55
103,057.49
105,751.03
108,679.17
110,594.11
113,867.3
114,175.79
115,641.28
115,816.21
116,983.8
118,546.05
118,671.85
121,597.55
122,616.84
122,782.75
124,153.04
127,056.21
14,681.4
35,673.41
42,559.73
800,000
600,000
400,000
200,000
0

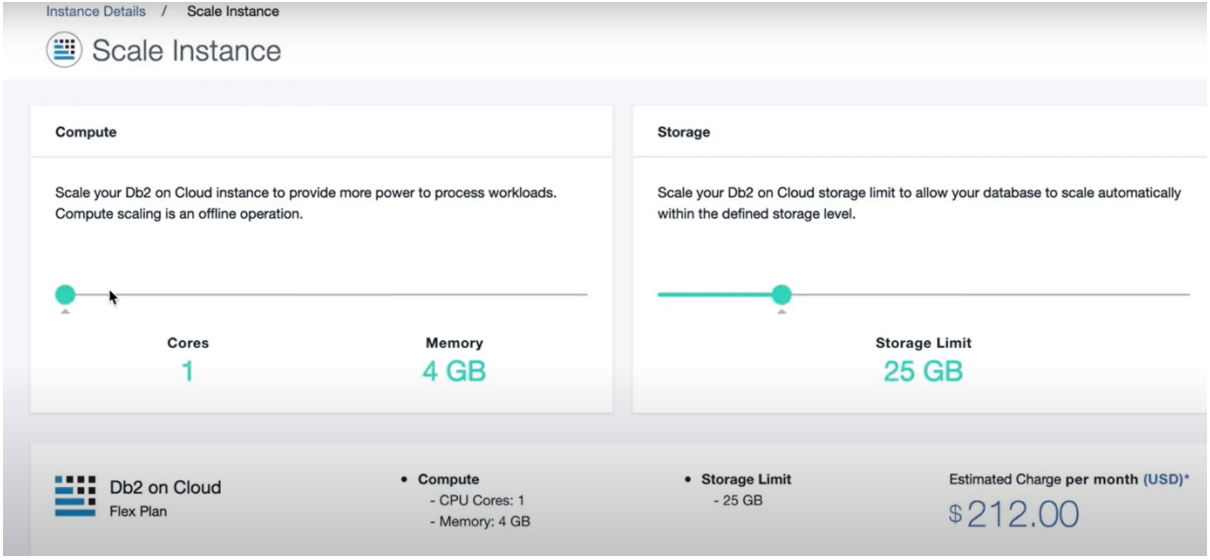
Data Responsiveness:



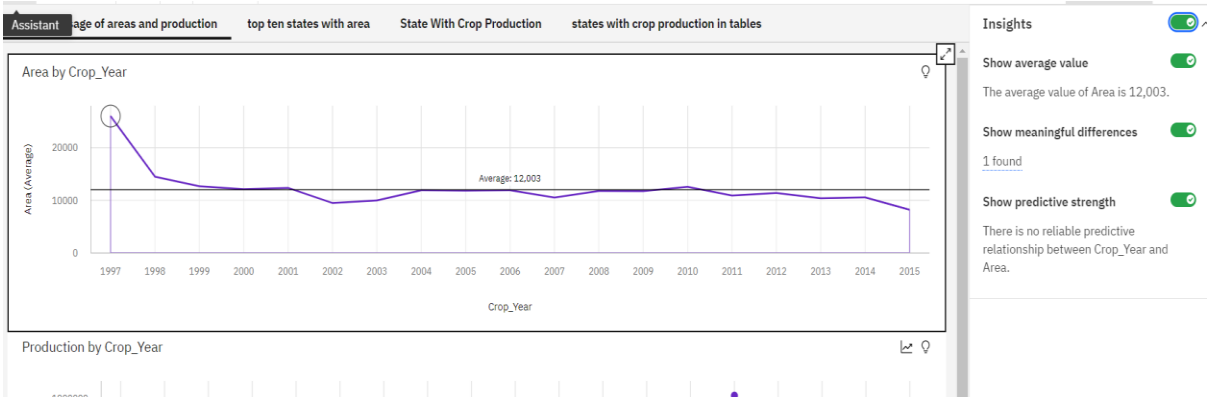
Prepare Empathy Map (Shruthi).docx	Add files via upload	2 months ago
Project Flow.pdf.docx	Add files via upload	2 months ago
README.md	Update README.md	2 months ago
Seasons With Average Productions.p...	Add files via upload	2 months ago
Top ten States with most area_dash b...	Add files via upload	2 months ago
Working With The Dataset assg 2.docx	Add files via upload	2 months ago
gopi python.ipynb	Add files via upload	last month

Amount Data to Rendered (DB2 Metrics)





Utilization of Data Filters



Descriptive Reports

Abstract: ***Analytics is the interpretation of data pattern that assist decision- making and performance improvement. Agriculture Data analytics in crop yield helps in analysing some important visualization, creating a dashboard and by going through these we will get most of the insights of Crop production in India. IBM Cognos Analytics integrates reporting, modelling, analysis, exploration, dashboards, stories, and event management so we can understand our organization's data, and make effective decisions. A dashboard helps us to monitor events or activities at a glance by providing key insights and analysis about our data on one or more pages or screens. In this project, we visualize, analyse and gain most of the insights by creating a dashboard.***
Keywords: IBM Cognos, Crop Production, Analytics and Agriculture.

I. INTRODUCTION

Agriculture is the backbone of Indian Economy. In India, majority of the farmers are not getting the expected crop yield due to several reasons. The agricultural yield is primarily depends on weather conditions. Rainfall conditions also influences the rice cultivation. In this context, the farmers necessarily requires a timely advice to predict the future crop productivity and an analysis is to be made in order to help the farmers to maximize the crop production in their crops. Yield prediction is an important agricultural problem. Every farmer is interested in knowing, how much yield he is about expect. In the past, yield prediction was performed by considering farmer's previous experience on a particular crop. The volume of data is enormous in Indian agriculture. The data when become information is highly useful for many purposes. IBM Cognos Business Intelligence is a web-based integrated business intelligence suite by IBM. It provides a toolset for reporting, analytics, score carding, and monitoring of events and metrics. The software consists of several components designed to meet the different information requirements in a company. IBM Cognos has components such as IBM Cognos Framework Manager, IBM Cognos Cube Designer, IBM Cognos Transformer. Cognos Analysis Studio helps business users get fast answers to business-related queries. Reporting studio allows you to create pixel-perfect reports for your organization. Cognos event studio allows you to assign a specific event that sends a notification to the stakeholder in your organization. Cognos Metric Studio allows you to monitor and analyze business metrics of your organization by building a scorecard environment.

II. LITERATURE SURVEY

M. A. Jayaram and Netra Marad, "Fuzzy interference Systems for Crop Prediction", Journal of Intelligent Systems, 2012, 21(4), pp.363-372[1]. Prediction of crop yield is significant in order to accurately meet market requirements and proper administration of agricultural activities directed towards enhancement in yield. Several parameters such as weather, pests, biophysical and morphological features merit their consideration while determining the yield. However, these parameters are uncertain in their nature, thus making the determined amount of yield to be approximate. It is exactly here that the fuzzy logic comes into play. This paper elaborates an attempt to develop fuzzy inference systems for crop yield prediction. Physio morphological features of Sorghum were considered. A huge database (around 1000 records) of physio morphological features such as days of 50 percent flowering, dead heart percentage, plant height, panicle length, panicle weight and number of primaries and the corresponding yield were considered for the development of the model. In order to find out the sensitivity of parameters, one-to-one, two-to-one and three-to-one combinations of input and output were considered. The results have clearly shown that panicle length contributes forth yield as the lone parameter with almost one-to-one matching between predicted yield and actual value while panicle length and panicle weight in combination seemed to play a decisive role in contributing for the yield with the prediction accuracy reflected by very low RMS value.

P. Vindya "Agricultural Analysis for Next Generation High Tech Farming in Data Mining" , Anna University, Trichy, Tamilnadu, India, 5 May 2015[2]. Recent developments in Information Technology for agriculture field have become an interesting research area to predict the crop yield [1].

In today's world, the amount of information stored has been enormously increasing day by day which is generally in the unstructured form and cannot be used for any processing to extract useful information using mining technique [2]. This paper presents a brief analysis of data mining methods and agriculture techniques, farm types, soil types, prediction using Multiple Linear Regression (MLR) technique for the selected region. This work mainly focuses on analyzing the agricultural analysis of organic farming and inorganic farming, time cultivation of the plant, profit and loss of the data and analyzes the real estate business land in a specific area and comparison of irrigated and unirrigated land. It concentrates organic, inorganic and real estate data sets from which the prediction in agriculture will be achieved. The purpose is to estimate difference in efficiency and prediction between organic and inorganic farming. This work aims at finding suitable data models that achieve a high accuracy and a high generality in terms of yield prediction capabilities.

Crop Yield Prediction Using Machine Learning A research group investigated the utilization of various information mining methods which will foresee rice crop yield for the data collected from the state of Maharashtra, India. A total of 27 regions of Maharashtra were selected for the assessment and the data was collected related to the principle rice crop yield influencing parameters such as different atmospheric conditions and various harvest parameters i.e Precipitation rate, minimum, average, maximum and most extreme temperature, reference trim cultivable area, evapotranspiration, and yield for the season between June to November referred as Kharif, for the years 1998 to 2002 from the open source, Indian Administration records. WEKA a Java based dialect programming for less challenging assistance with information data sets, assigning design outcomes tool was applied for dataset processing and the overall methodology of the study includes, (1) pre-processing of dataset (2) Building the prediction model utilizing WEKA and (3) Analyzing the outcomes. Cross validation study is carried out to scrutinize how a predictable information mining method will execute on an ambiguous dataset. Study applied 10-fold higher cross validation study design to assess the data subsets for screening and testing. Identified and collected information was randomly distributed into 10 sections where in one data section was used for testing while all other data sections were utilized for the preparation information. Study reported that the method applied was supportive in the precise estimation of rice crop yield for the state of Maharashtra, India. The precise quantification of the rice productivity in various climatic conditions can help farmer to understand the optimum condition for the higher rice crop yield [8].

Simulation models based on field experiment are valuable technologies for studying and understanding crop yield gaps, but one of the critical challenge remain with these methods is scaling up of these approach to assess the data collated between different time intervals from the broader geographical regions. Satellite retrieved data have frequently been revealed to present data sets that, by itself or in grouping with other information and model designs, can precisely determine the yields of crop in agricultural lands. The yield maps developed shall provide an unique opportunity to overcome both spatial and temporal based scaling up challenges and thus improve the ideology of crop yield gaps prediction. A review was conducted to discuss the applications of remote sensing technology to determine the impact and causes of yield gaps. Even though the example discussed by the research group demonstrates the usefulness of remote sensing in the prediction of yield gaps, but also many areas of possible application with respect to the crop yield assessment, prediction and improvement remain unexplored. Study proposed two less complicated, easily assessable methods to determine and quantify the yield gaps between various agricultural fields. First method works closely with the constructive maps representing the average crop yields, it can be used directly to accesses specific crop yield influencing factors for further studies whereas the second method use the remote sensing technology to retrieve the data for providing the useful information regarding the crop yield prediction and estimation [14].

III. SYSTEM DESIGN

System design thought as the application of theory of the systems for the development of the project. System design defines the architecture, data flow, use case, class, sequence and activity diagrams of the project development.

A. IBM Cognos Analytics

IBM Cognos Analytics is a set of business intelligence tools available on cloud or on- premise. The primary focus is in the area of Descriptive Analytics, to help users see the information in your data through dashboards, professional reporting and self-service data exploration. In this work, we used the IBM cognos data analytics for analysing the crop yield data.

Following are important features of IBM Cognos:

1) *Get Connected* - Connect your data effortlessly Import data from CSV files and spreadsheets. Connect to cloud or on-premises data sources, including SQL databases, Google BigQuery, Amazon, Redshift, and more.
2) *Prepare your data* – Prepare and connect data automatically Save time cleaning your data with AI-assisted data preparation.