



# Data Science in Action using Dataiku

An AAI Artificial Intelligence– Technical Track Course



## Course Outline





# Course Outline

1. Introduction to Course

2. Set up sandbox

## Install Dataiku on Mac

Dataiku can be installed directly on Mac OS 10.9 and later. Follow the steps below.



## Install Dataiku using Virtualbox Or VMWare

Dataiku provides a pre-built Linux virtual machine for the free Virtualbox engine. This lets you run DSS for evaluation purpose on Windows or Mac OS

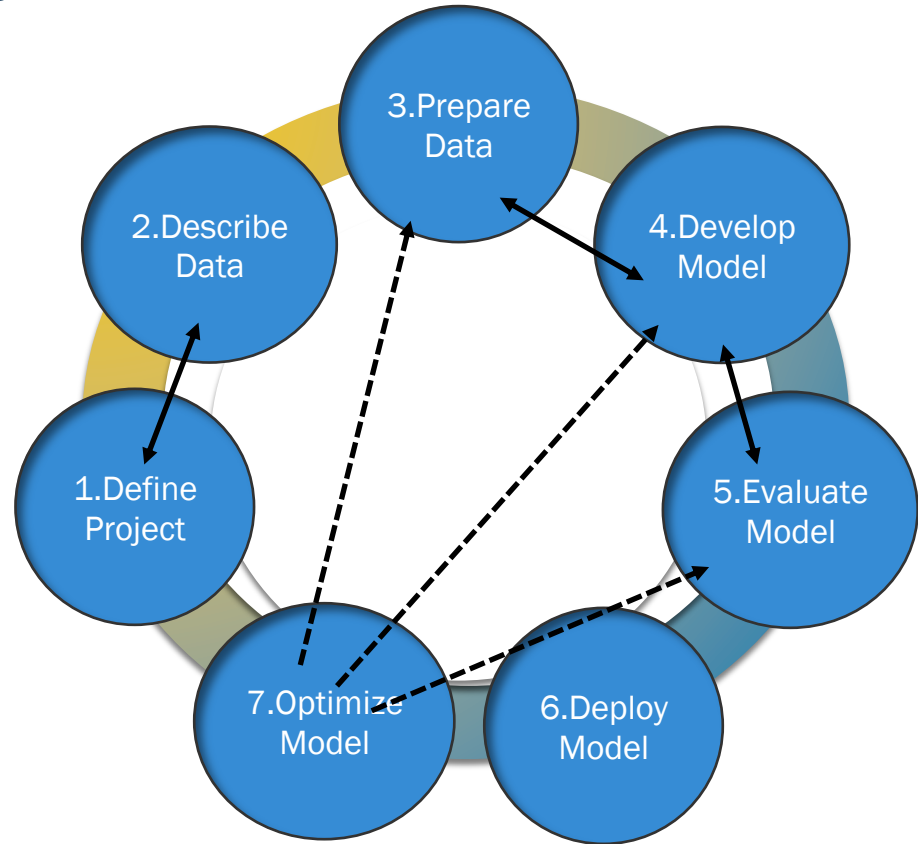


## Microsoft Azure: AI-Driven Projects at Scale with Dataiku



# Course Outline

1. Introduction to Course
2. Set up sandbox
3. Data Science Methodology





# Course Outline

1. Introduction to Course
2. Set up sandbox
3. Data Science Methodology
4. Step 1 – Define Project

User Persona



Needs / Benefits



Goals / Success Criteria



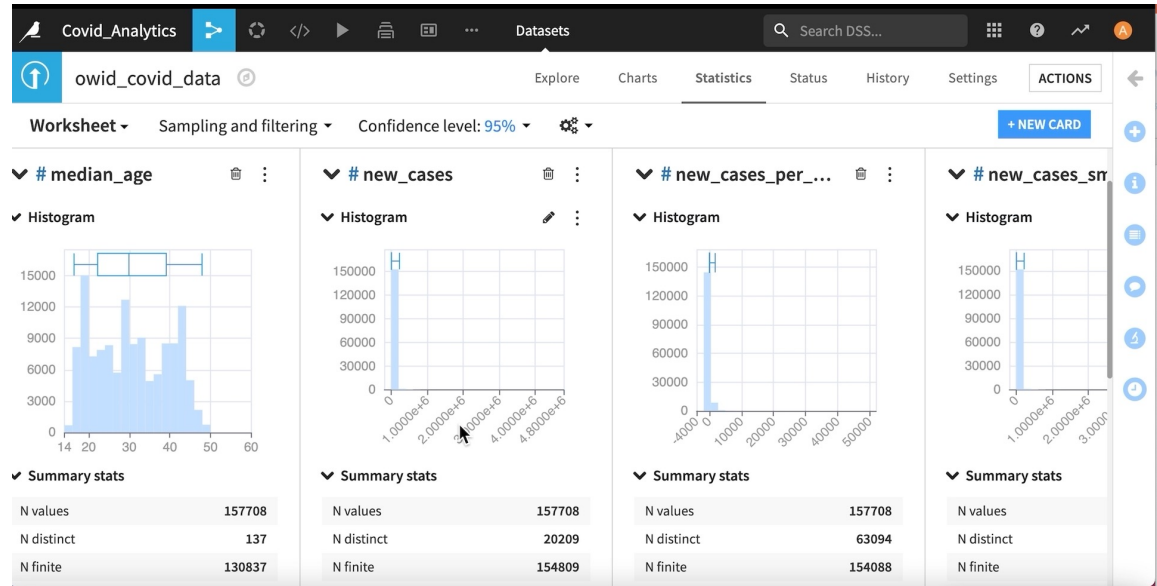
Project Plan





# Course Outline

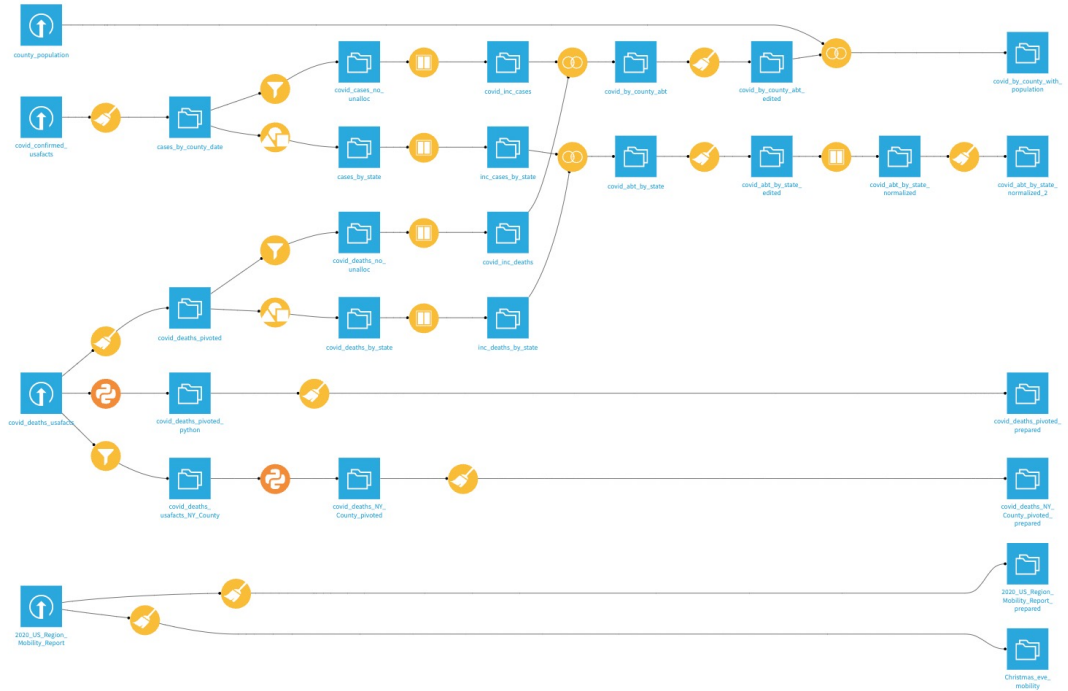
1. Introduction to Course
2. Set up sandbox
3. Data Science Methodology
4. Step 1 - Define Project
5. Step 2 – Describe Data





# Course Outline

1. Introduction to Course
2. Set up sandbox
3. Data Science Methodology
4. Step 1 - Define Project
5. Step 2 - Describe Data
- 6. Step 3 - Prepare Data**

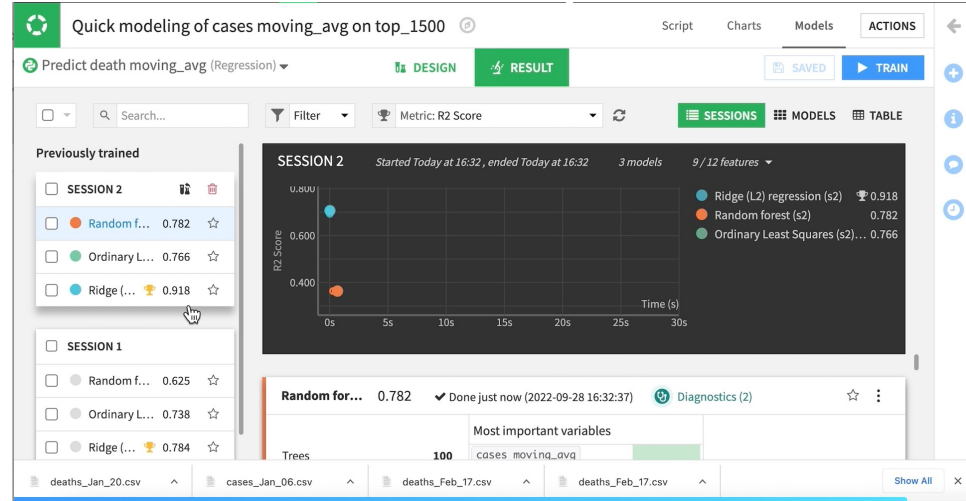




# Course Outline

1. Introduction to Course
2. Set up sandbox
3. Data Science Methodology
4. Step 1 - Define Project
5. Step 2 - Describe Data
6. Step 3 - Prepare Data
- 7. Step 4 – Develop Model**

Random Forest	<input checked="" type="checkbox"/>
Gradient tree boosting	<input type="checkbox"/>
Ordinary Least Squares	<input checked="" type="checkbox"/>
Ridge Regression	<input checked="" type="checkbox"/>
Lasso Regression	<input type="checkbox"/>
LightGBM	<input type="checkbox"/>
XGBoost	<input type="checkbox"/>
Decision Tree	<input type="checkbox"/>



Support Vector Mach...	<input type="checkbox"/>
Stochastic Gradient ...	<input type="checkbox"/>
KNN	<input type="checkbox"/>
Extra Random Trees	<input type="checkbox"/>
Neural Network	<input type="checkbox"/>
Lasso Path	<input type="checkbox"/>
+ ADD CUSTOM PYTHON	



# Course Outline

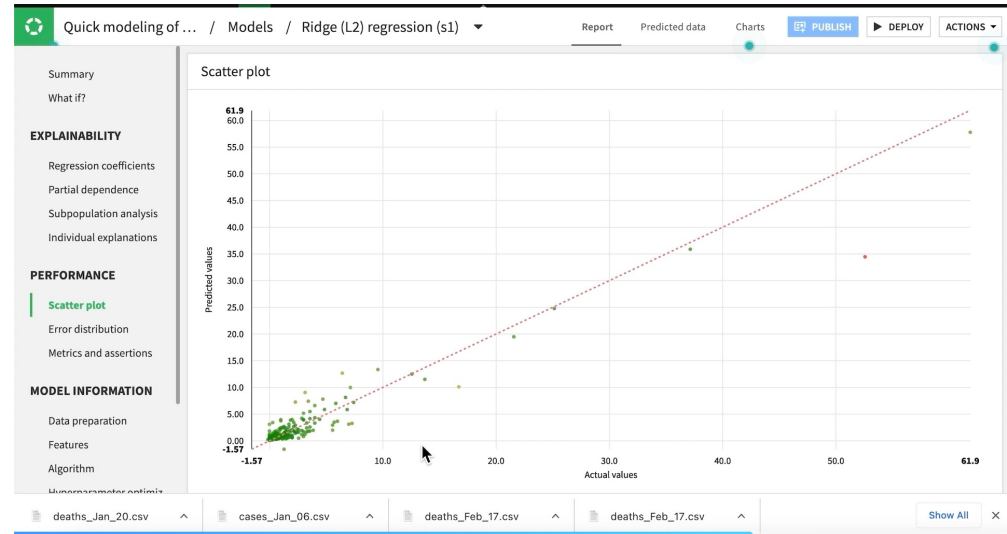
1. Introduction to Course
2. Set up sandbox
3. Data Science Methodology
4. Step 1 - Define Project
5. Step 2 - Describe Data
6. Step 3 - Prepare Data
7. Step 4 - Develop Model
8. Step 5 - Evaluate Model

Ridge (L2) regression (s1)

R2 Score: 0.918

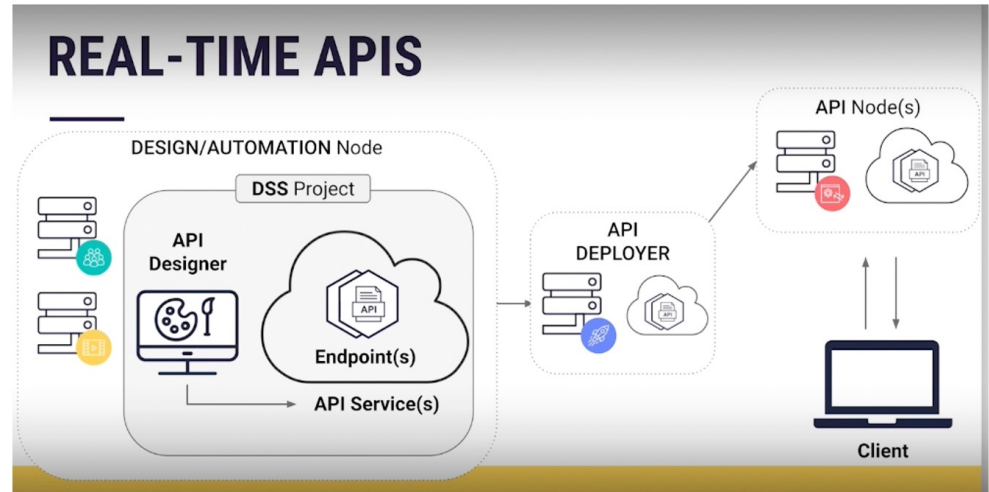
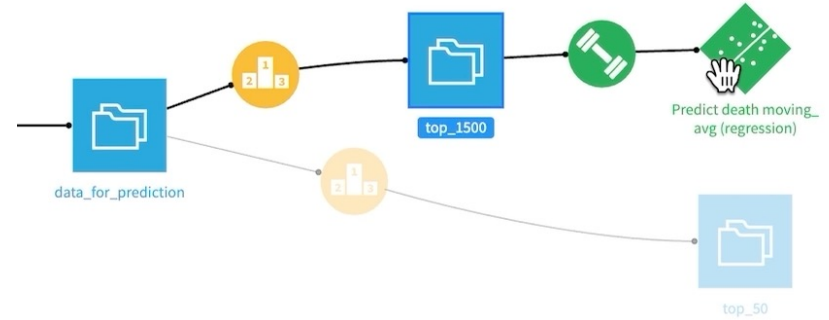
## Model

Model ID	A-COVID_ANALYSIS_PREDICTION-AOpja1Kq-xruvPC39-s1-pp2-m1
Backend	Python (in memory)
Algorithm	Ridge regression
Trained on	2022/09/28 18:04
Columns	12
Train set rows	1211
Test set rows	289
Calibration method	No calibration





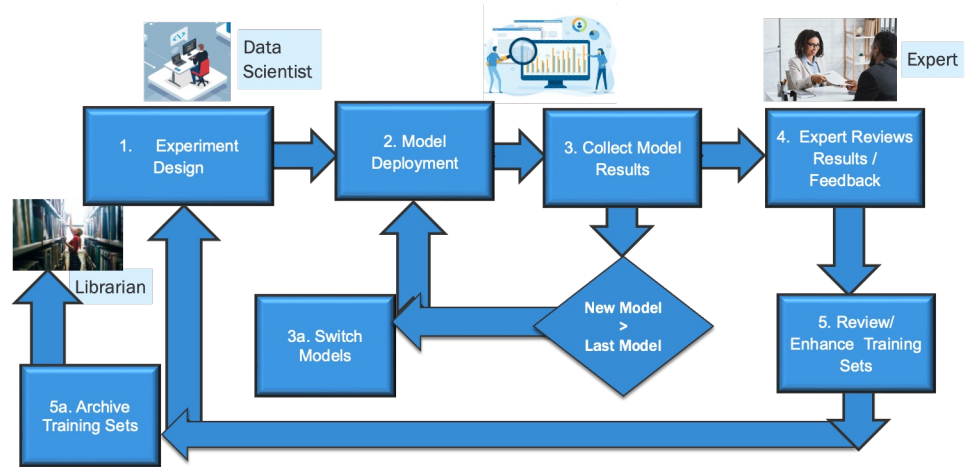
1. Introduction to Course
2. Set up sandbox
3. Review Data Science Methodology
4. Step 1 – Define Project
5. Step 2 – Describe Data
6. Step 3 - Prepare Data
7. Step 4 – Develop Model
8. Step 5 – Evaluate Model
9. Step 6 - Deploy Model





# Course Outline

1. Introduction to Course
2. Set up sandbox
3. Review Data Science Methodology
4. Step 1 – Define Project
5. Step 2 – Describe Data
6. Step 3 - Prepare Data
7. Step 4 – Develop Model
8. Step 5 – Evaluate Model
9. Step 6 - Deploy Model
10. Step 7 - Optimize Model



1. Introduction to Course
2. Set up sandbox
3. Review Data Science Methodology
4. Step 1 – Define Project
5. Step 2 – Describe Data
6. Step 3 - Prepare Data
7. Step 4 – Develop Model
8. Step 5 – Evaluate Model
9. Step 6 - Deploy Model
10. Step 7 – Optimize Model
11. Summary and Next Steps

