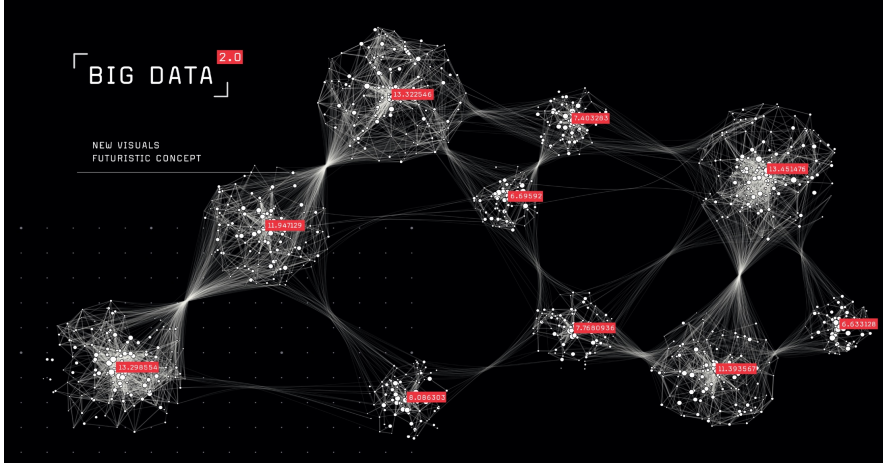




Data Science in Action using ikigailabs

An AAI Artificial Intelligence– Technical Track Course



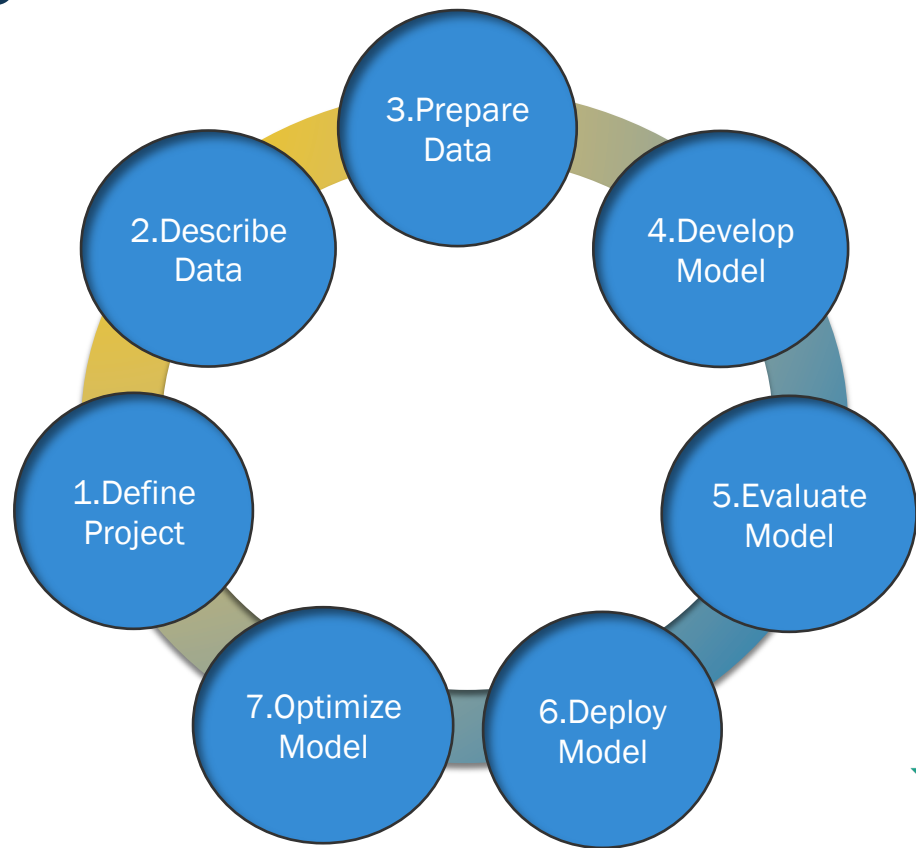
Course Outline





Course Outline

1. Introduction to Course



ikigai



Course Outline

1. Introduction to Course

2. Set up sandbox

1. Open the Ikigai platform

<https://app.ikigailabs.io/>

ikigai

Start your 14-day trial of Ikigai for free

Take advantage of the full spectrum of Ikigai's platform.
No credit card or data science knowledge required.

- Connect and stitch together data from any sources in just a few easy steps
- Use pre-built solutions or create your custom data apps
- Leverage powerful machine learning models to run what-if simulations, trade-off analysis, forecasting and predictive analytics with no code
- Operationalize your insights by taking actions directly from the dashboards

Copyright © IkigaiLabs.io 2023.

Create Account
To get started, create a new user account

neena.sathi@gmail.com

Password*

Confirm*

I have read and agree to the terms of [Ikigai Labs License Agreement](#)

Get Started

Already have an account? [Sign in](#)



Course Outline

1. Introduction to Course
2. Set up sandbox
3. Step 1 – Define Project



Define Project – Health Equity Use Case Example

User Persona



Citizens



Healthcare Administration Staff

Needs / Benefits



Objective:

Need to understand current health infrastructure to expand services in new communities

Queries:

- What are social and health profile of a specific communities in the state of California?
- How do various counties compare to each other based upon their social, demographic and health resources

Goals / Success Criteria



Can we make health equity score for communities based upon its

- Demographic, social and health resources
- 90% Accuracy
- Time boxed

Project Plan



Tasks

Start	W 1	W 2	W 3	W 4
Project Planning				
Data Collection				
Data Preparation				
Modeling				
Visualization				
Testing				
Deployment				

Deliverables



Plan Model Dashboard

Resources

- 2 resources
- Cloud Sandbox



ikigai

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

1. Open Ikigai Labs
2. Select new project
3. Describe your project
4. Click 'Create' button
5. Click on the "<" button at the top left corner
6. Select Datasets and click + icon
7. Select "Upload Dataset" option
8. Provide Dataset Name
9. Select the file to be uploaded.
10. Click Upload.

Health Equity Project

- Flows
- Datasets
- Connectors
- Databases
- Dashboards
- Models
- Resources

Upload Dataset

No Datasets Four

Upload New Dataset

Dataset name* 0 / 40
Enter new name for Dataset

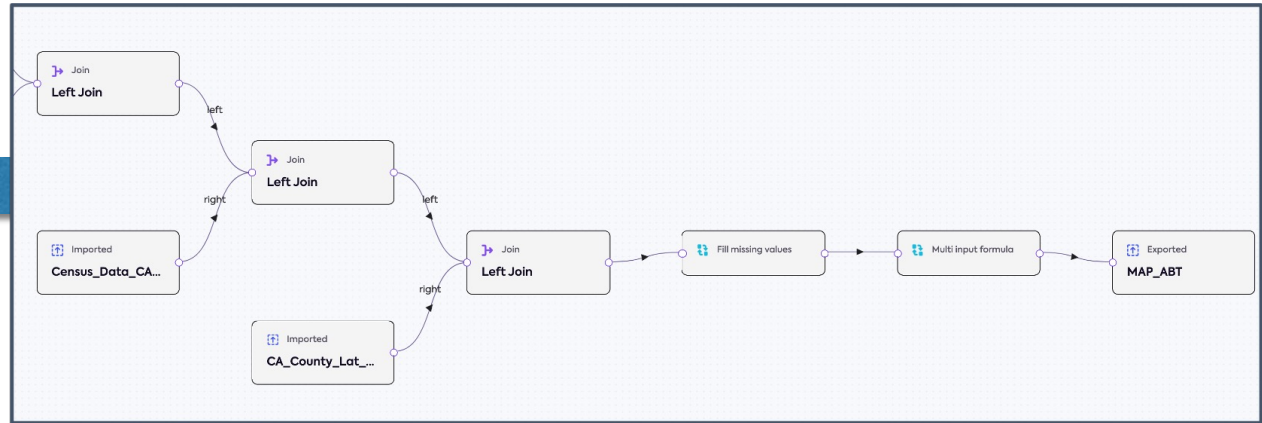
Folder name
Type to search Folder Name

Drop the file or click to select file.

Cancel Upload



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





Course Outline

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

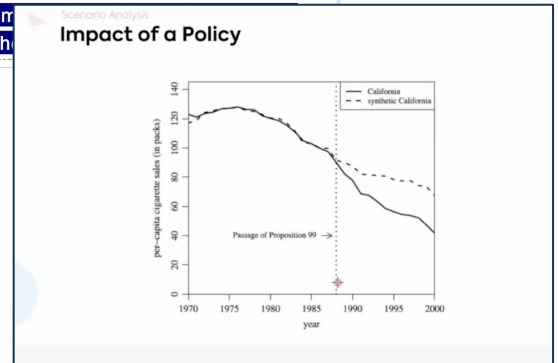


K-Means Clustering Results

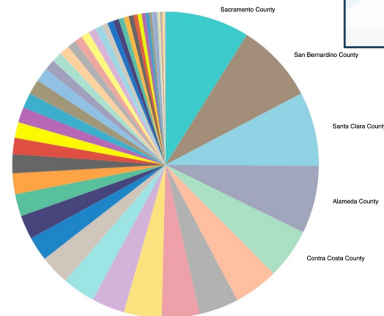
Cluster_ID	Label	# of Counties	Number of Enrollments
0	Central Valley Counties	10	2,999
1	Urbanized Coastal and Adjacent Counties	21	9,881
2	Rural Interior Counties	6	173
3	Megacity County	1	10,986
4	Rural Northern and Inland Counties	20	505

Rural interior counties have least health

enroll
most h



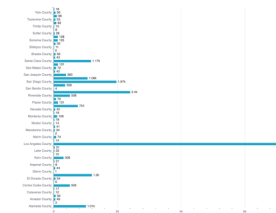
Cancer Total – Pie chart



Top 5 Counties

CountyFullName	Cancer_Rate_per_100K
Lake County	78.2
Glenn County	77.4
Amador County	73.5
Colusa County	68.3
Tehama County	67.7

Health Vulnerability
• Total Enrollments – Bar chart



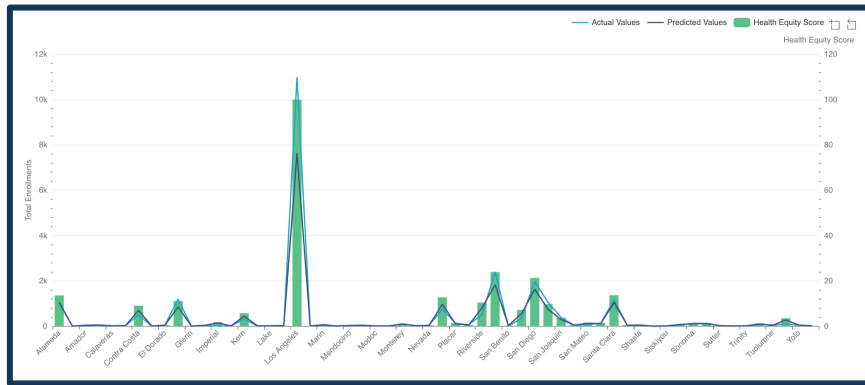


Course Outline

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

County	Total Enrollments	Cluster_ID	Cluster Score
Alameda	1011	1	0.01
Alpine	3	4	0.65
Amador	49	4	0.33
Butte	59	4	0.17
Calaveras	12	2	0.76
Colusa	17	4	0.70
Contra Costa	506	1	0.21
Del Norte	6	4	0.45
El Dorado	54	1	0.91

- Counties with high cluster scores such as El Dorado, Calaveras are well inside their corresponding cluster zone and close to the centroid.
- Counties with low score such as Alameda, Butte fall at the overlapping region of neighboring clusters and are away from the centroid of the cluster.



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

4. GET the prediction output results

← → C app.ikigai labs.io/dataset/2TcPXlryS6AHz5zVF0T1BkGreXL

```
dataset_id = "2TcPXlryS6AHz5zVF0T1BkGreXL"
url = f"https://api.ikigai labs.io/pypr/get-dataset-download-url?dataset_id={dataset_id}"

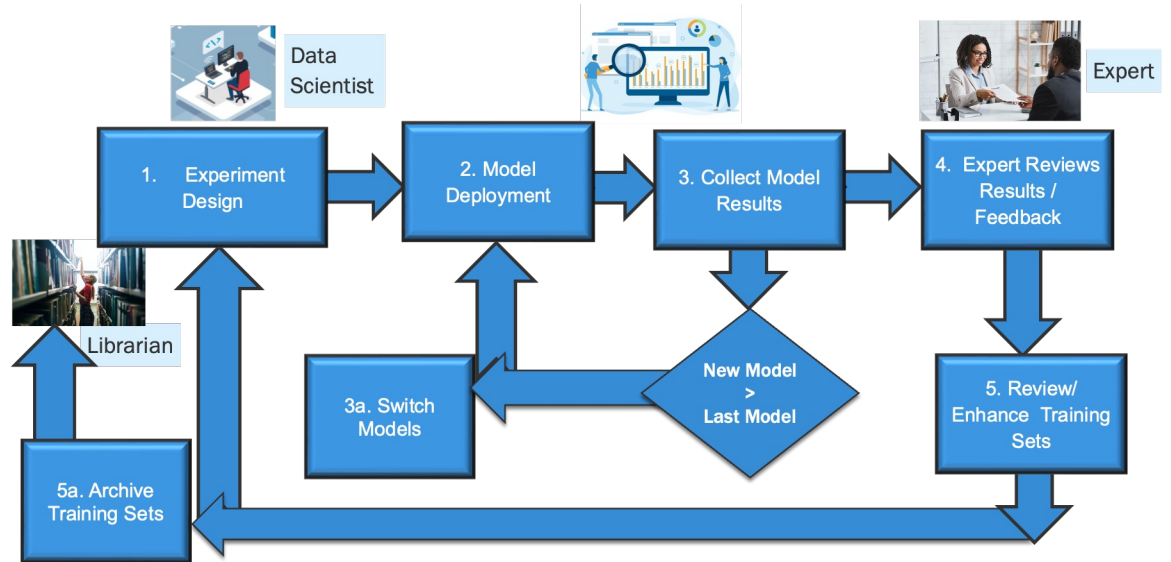
# We send an empty payload and the same security configuration in the headers
payload = {}
headers = {
    'User': user_email,
    'Api-key': api_key,
    'Content-type': 'application/json',
}
```

Get the prediction results output:

- Open the csv file in Ikgai to get the dataset id from the `url`.
- Use the GET `api` call to retrieve data from the `api-end point` https://api.ikigai labs.io/pypr/get-dataset-download-url?dataset_id={dataset_id}
- The API response would be the prediction results in raw `json` format.



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



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

