(name of your school)

CAPSTONE PROJECT

Project Documentation

“DISTRICTS WITH SIMILAR HEALTH NEEDS”

SDG 3: Good Health and Well-being · SDG 10: Reduced Inequalities

Class XII

Artificial Intelligence — Subject Code 843
Part C: Capstone Project

Session _____

Submitted by:

Group members listed overleaf

Under the guidance of:

_____

Teacher (Artificial Intelligence)

GROUP MEMBERS

The Capstone Project is carried out by a group of three to five students.

S. NO.	NAME	ROLL NO.	ROLE IN THE PROJECT	SIGNATURE
1				
2				
3				
4				
5				

Certificate

This is to certify that the students named above, of Class XII, have satisfactorily completed the Capstone Project entitled **"Districts with Similar Health Needs"** during the academic session _____, as prescribed in Part C of Artificial Intelligence (Subject Code 843) by the Central Board of Secondary Education, and have submitted this documentation and the accompanying three-minute video.

Teacher-in-charge

External Examiner

Principal

ACKNOWLEDGEMENT

We would like to thank _____, our Artificial Intelligence teacher, for guiding this capstone through every step of the Data Science Methodology.

We are grateful to the Principal of _____ for the computer lab and the support that made the project possible.

Our thanks go to _____, who helped us understand the problem and how it is handled today, and to everyone who shared data with us or answered our questions.

Finally, we thank our families for their patience while the project was built, tested and filmed.

_____
(Group members)

INDEX

SL. NO.	TOPIC	PAGE NO.
1	Introduction	5
2	Business Understanding	6
3	Analytic Approach	6
4	Data Requirements	8
5	Data Collection	8
6	Data Understanding	10
7	Data Preparation	10
8	Modelling	12
9	Evaluation	12
10	Deployment	14
11	Feedback	14
12	Ethical Considerations	14
13	The Project in Orange	14
14	Technical Details	15
15	Source Code	15
16	Output of a Real Run	18
17	Testing	19
18	Team Roles and Timeline	20
19	Results and Future Scope	20
20	Conclusion	21
21	Video Script	21
22	Viva Preparation	21
23	Glossary	21
24	References	22
25	Appendix: The Data	22

INTRODUCTION

Surveys such as the National Family Health Survey (NFHS) report dozens of health indicators for every district: how many births happen in hospitals, how many children are fully immunised or stunted, how many women are anaemic, and more.

A state with dozens of districts and dozens of indicators cannot plan district by district and indicator by indicator — but a single plan for the whole state ignores how different districts are. Some need hospitals and immunisation drives; others already have those, but children are still undernourished.

This capstone groups districts that look alike across eight indicators, using K-means clustering. Nobody has labelled the districts, so this is unsupervised learning: the method must find the groups on its own, and the team must decide how many groups there are and what each one means.

Who would use it: The state health department and the National Health Mission's state and district teams, who plan programmes and share out funds.

Alignment to the Sustainable Development Goals

SDG 3 — Good Health and Well-being; SDG 10 — Reduced Inequalities. SDG 3 asks for healthy lives for all at all ages — target 3.1 on safe births, 3.2 on child deaths, 3.8 on health coverage. SDG target 10.2 asks for the inclusion of all, irrespective of where they live. Within one state, some districts are far behind others. Finding which districts share the same needs lets the state aim the right programme at the right places, so that the districts furthest behind are not served by a plan made for the average.

The Data Science Methodology

The project follows the ten-step Data Science Methodology of John Rollins (IBM), taught in Unit 2 of the Class XII handbook. The sections that follow take the steps in order.

MODULE	STEPS
1. From Problem to Approach	Business Understanding; Analytic Approach
2. From Requirements to Collection	Data Requirements; Data Collection
3. From Understanding to Preparation	Data Understanding; Data Preparation
4. From Modelling to Evaluation	Modelling; Evaluation
5. From Deployment to Feedback	Deployment; Feedback

BUSINESS UNDERSTANDING

What is the problem that you are trying to solve?

The problem was scoped with the 5W1H Problem Canvas.

Who? • Families in districts that are behind • The state health department • District health officers	**What?** • One plan for all districts misses their different needs • Too many indicators to compare by eye
Where? • All the districts of one state	**When?** • Each time a new survey round is published, when plans are made
Why? • Programmes and funds reach the districts that need them • Each programme fits the need it is meant for	**How?** • Cluster districts on eight survey indicators • Describe each group by its weakest indicators, and plan for each group

Problem statement

> Our **the state health department** have a problem of **matching health programmes to what each district needs** when **districts differ widely, and each survey gives dozens of indicators for each of them**. An ideal solution would **a few groups of similar districts, each with a clear description of what it needs most**.

Talking to the stakeholders

The handbook pairs the 5W1H canvas with the Design Thinking framework, which starts by empathising with the people who have the problem. These are the questions the group put to them, and what each answer changed in the project.

QUESTION ASKED	WHAT WE LEARNED, AND WHAT IT CHANGED
To a state health planner: how are programmes planned now?	Mostly state-wide, with extra attention to a list of the 'worst' districts. But 'worst' at what? That question led to clustering instead of ranking.
To the same planner: how many groups could you plan for?	Three to six. One plan per district is too many; two groups would be too crude. That set the range of k to try.
To a district health officer: which indicators matter most?	Births in hospitals, immunisation, child stunting and anaemia — and the things around them: sanitation, clean cooking fuel, women's literacy and health insurance. Those became the eight indicators.
To the officer: is the survey data clean?	Mostly, but numbers are typed in by hand — a missing decimal point is common. So values above 100 percent are checked.
To a public health teacher: why not simply rank the districts?	A rank gives one number. Two districts can have the same rank for different reasons — one lacks hospitals, the other has hospitals but hungry children. They need different programmes.
To the planner: what will you do with the groups?	Use each group's weak points to choose its programme — and check after the next survey whether districts have moved. That shaped deployment and feedback.

How the problem is handled today

State-wide plans. Simple to run, but treat a district with no hospital and one with hungry children in the same way.

Ranking districts. Lists of the furthest-behind districts, such as the Aspirational Districts Programme's composite index, focus attention well — but a rank is one number, and hides why a district is behind.

Reading every indicator. Officers can study each district's factsheet, but no one can compare hundreds of numbers across dozens of districts by eye.

Objectives

1. To frame district health planning with the 5W1H canvas and interviews
2. To prepare eight district indicators for clustering: duplicates, typing errors, blanks and scaling
3. To cluster districts with K-means, choosing the number of groups with the elbow method and the silhouette score
4. To check that the groups are stable and describe each group by its weakest indicators
5. To place a new district in a group, and use the next survey as feedback

ANALYTIC APPROACH

How can we use data to answer the question?

Type of analytics: Descriptive.

The question is not to predict anything but to describe the districts: which ones are alike, and how. That is descriptive analytics.

There are no labels — no one has said which group each district belongs to — so the model must be unsupervised learning, and the technique is clustering.

Approaches considered

CANDIDATE	WHY IT WAS CONSIDERED
K-means clustering	Places k centres, puts each district with its nearest centre, moves each centre to the middle of its districts, and repeats until nothing changes.
Elbow method	Plots how tightly packed the groups are (inertia) against k; the bend in the curve suggests a good k.
Silhouette score	For each district, compares how close it is to its own group with how close it is to the next group. The average, from −1 to 1, measures how clearly separated the groups are.

DATA REQUIREMENTS

What are the data requirements?

The data needs one row per district and the same indicators for every district, from the same survey round, so that districts are compared fairly.

The indicators must describe needs that programmes can act on. Eight were chosen from the interviews: four health outcomes and services, and four conditions around them.

DATA NEEDED	TYPE	FORMAT	SOURCE
Births in a health facility	Percent	of births in the last 5 years	NFHS district factsheet
Children fully immunised	Percent	of children aged 12–23 months	NFHS district factsheet
Children stunted	Percent	of children under 5 (lower is better)	NFHS district factsheet
Women anaemic	Percent	of women aged 15–49 (lower is better)	NFHS district factsheet
Improved sanitation, clean fuel	Percent	of households	NFHS district factsheet
Women literate, health insurance	Percent	of women / of households	NFHS district factsheet

DATA COLLECTION

What occurs during data collection?

In a real project, these indicators come from the NFHS district factsheets, which the Ministry of Health and Family Welfare and the International Institute for Population Sciences publish for every district. That is secondary data; the interviews with planners and officers are primary data.

The sample file, districts.csv, holds 301 rows for 300 imaginary districts, with 8 indicators each. districts_next_survey.csv holds the same districts in a later survey round, used for feedback. The data is a LambdaLab sample: realistic in range and pattern, but not real district figures.

> **About this dataset.** The data is a LambdaLab sample made by a program to behave like the real thing, so the project could be run and tested. It is not real data. A submitted capstone must use data the group collected or a credited public dataset.

The columns of the dataset

COLUMN	TYPE	WHAT IT HOLDS	ROLE
`district`	text	District name	identifier, not used for clustering
`institutional_births_pct`	decimal	Births in a hospital or health centre, %	feature
`full_immunisation_pct`	decimal	Children fully immunised, %; some blank	feature
`stunted_children_pct`	decimal	Children too short for their age, % (lower is better)	feature
`anaemic_women_pct`	decimal	Women with anaemia, % (lower is better)	feature
`improved_sanitation_pct`	decimal	Households with improved sanitation, %	feature
`clean_fuel_pct`	decimal	Households cooking with clean fuel, %; some blank	feature
`women_literacy_pct`	decimal	Women who can read and write, %	feature
`health_insurance_pct`	decimal	Households with health insurance, %; some blank	feature

districts.csv — Eight health indicators for each district. 301 rows, 15.7 KB.

districts_next_survey.csv — The same districts in the next survey, used only for feedback. 300 rows, 15.7 KB.

The first eight rows of districts.csv

DISTRICT	INSTITUTIONAL_BIRTHS_PCT	FULL_IMMUNISATION_PCT	STUNTED_CHILDREN_PCT	ANAEMIC_WOMEN_PCT	IMPROVED_SANITATION_PCT	CLEA
District 101	93.5	81.1	34.4	54.9	72.7	37.7
District 102	63.9	59.4	40.1	57.0	51.5	16.8
District 103	83.9	71.5	32.0	43.4	92.9	67.4
District 104	98.8	93.1	18.4	52.1	93.8	73.5
District 105	84.6	64.4	31.9	54.0	84.6	55.6
District 106	65.3	50.3	45.8	69.0	50.2	25.8
District 107	65.6	52.2	45.2	65.2	58.2	
District 108	95.1	86.5	13.5	48.6	90.3	82.8

The first four rows of districts_next_survey.csv

DISTRICT	INSTITUTIONAL_BIRTHS_PCT	FULL_IMMUNISATION_PCT	STUNTED_CHILDREN_PCT	ANAEMIC_WOMEN_PCT	IMPROVED_SANITATION_PCT	CLEA
District 101	94.9	85.6	28.6	55.7	78.0	42.1
District 102	66.4	62.9	39.5	56.0	55.4	21.6
District 103	86.1	76.6	29.7	38.7	95.9	73.7
District 104	99.5	99.5	15.0	52.8	95.3	78.8

DATA UNDERSTANDING

Is the data collected representative of the problem to be solved?

The first checks find blank cells in three indicators and values above 100 percent — impossible for a percentage, and a sign of a missing decimal point.

Chart 1 shows how the indicators move together. Five of them move strongly as a block: districts with more immunisation, clean cooking fuel and health insurance tend to have less stunting and anaemia (correlations of 0.5 to 0.8). Hospital births, sanitation and literacy move with them only weakly. So districts can be ahead on some things and behind on others — which is exactly why grouping them is more useful than ranking them.

Chart 2 shows that the indicators do not vary equally: clean cooking fuel varies the most across districts, stunting and anaemia the least. This matters for K-means, which measures distances.

Figure 1: How the indicators move together

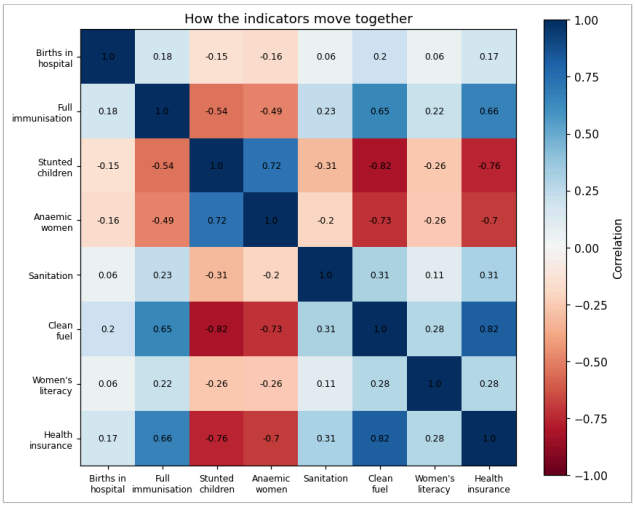


Figure 1 — How the indicators move together, a heat map drawn by the program (chart1_correlation.png).

How to read it. The correlation between every pair of indicators: blue for moving together, red for moving in opposite directions.

Immunisation, clean fuel and insurance move together (blue), and stunting and anaemia move against them (red) — a strong block. Hospital births, sanitation and literacy are only weakly linked to the rest, so a district can be ahead on those and behind on nutrition, or the reverse.

DATA PREPARATION

What additional work is required to manipulate and work with the data?

PROBLEM	HOW MUCH	WHAT WAS DONE, AND WHY
A district entered twice	`1`	Removed by district name, so no district counts twice in its group.
Percentages typed without the decimal point	`3`	Any value above 100 is divided by 10 — 912 becomes 91.2 — as the officer described.
Blank cells	`12`	Filled with that indicator's median across districts. K-means cannot work with blanks, and dropping the districts would leave them out of every group.
Indicators with different spreads	`Clean fuel varies about two and a half times as much as stunting`	Every indicator is rescaled to mean 0 and standard deviation 1 with StandardScaler, so each counts equally in the distances.

Feature engineering

No new features are made: each district is described by its eight indicators. What matters is scaling. Without it, a 20-point difference in clean fuel and a 20-point difference in stunting would count the same, though 20 points is ordinary for clean fuel and huge for stunting. After scaling, differences are measured in standard deviations, which is fair to every indicator.

Two indicators are 'lower is better' (stunting and anaemia). K-means does not care about direction, but describing the groups does, so the program flips their sign when it judges whether a group is better or worse than the state average.

Figure 2: The spread of each indicator

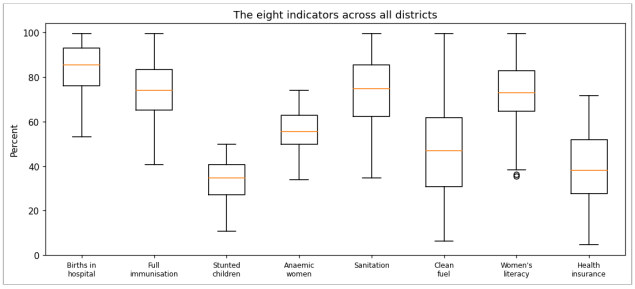


Figure 2 — The spread of each indicator, a box plot drawn by the program (chart2_spread.png).

How to read it. A box plot of each indicator across all districts.

Clean fuel spreads across almost the whole range; stunting and anaemia sit in narrow bands. Without scaling, clean fuel would dominate the distances K-means measures.

MODELLING

In what way can the data be visualised to get to the answer that is required?

AI model ▶ Learning based ▶ Unsupervised learning ▶ Clustering

There are no labels to learn from, so the districts cannot be split into training and testing sets to check predictions. Instead, the model is judged by how clearly separated and how stable its groups are, and by whether the groups make sense to the people who will use them.

K-means needs k, the number of groups, decided in advance. The planner said three to six groups could be used; the program tries k from 2 to 8 and uses the elbow method and the silhouette score to choose.

How the models work

For each k, K-means is run with n_init=10: it starts from ten different random positions and keeps the best result, because a single start can end in a poor grouping. The silhouette was highest at 0.4115, for k = 3, with k = 4 almost equal; the elbow is at 4. When two values of k have almost the same silhouette, the program keeps the larger, because more groups mean more targeted programmes. Number of groups chosen: 4.

K-means numbers its clusters in no particular order, so the program renames them by need: each group's centre is averaged across the eight indicators (with stunting and anaemia flipped), and Group 1 is the group that is worst off.

Figure 3: Choosing the number of groups

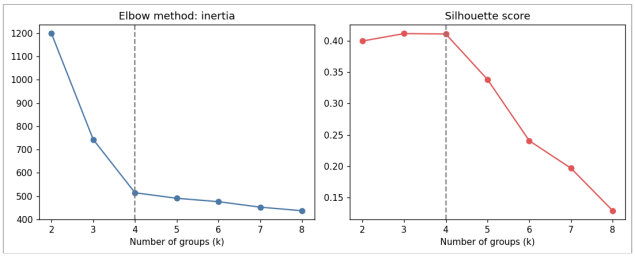


Figure 3 — Choosing the number of groups, a line chart drawn by the program (chart3_elbow_silhouette.png).

How to read it. Left: inertia (how tightly packed the groups are) for k from 2 to 8. Right: the silhouette score for each k. The dashed line marks the k chosen.

Inertia falls steeply up to k = 4 and then flattens — the elbow is at 4. The silhouette is highest for 3 and 4, almost equally, and falls quickly after 4. Both point to four groups.

EVALUATION

Does the model used really answer the initial question, or does it need to be adjusted?

The silhouette score of the four groups is 0.411, which means the groups are reasonably well separated, though not perfectly — real districts form a range, not islands. 0 districts had a negative silhouette, which would mean being closer to another group than to their own.

K-means depends on its starting point, so the clustering was repeated from ten other random starts. The agreement with the original grouping, measured by the adjusted Rand index (1 means identical), was at lowest 1.0. The groups are a property of the data, not of luck.

The groups have 70, 84, 86, 60 districts. The most important check is whether they mean something. Each group's weakest indicators compared with the state average are: Group 1 — births in hospital, full immunisation; Group 2 — stunted children, anaemic women; Group 3 — full immunisation; Group 4 — none - above average on all.

This is what a ranking would have hidden. Groups 2 and 3 are both in the middle of the state, but for opposite reasons: Group 2 has the hospitals and the vaccines but undernourished children and women; Group 3 has better nutrition and sanitation but lags on immunisation. They need different programmes.

Figure 4: What each group looks like

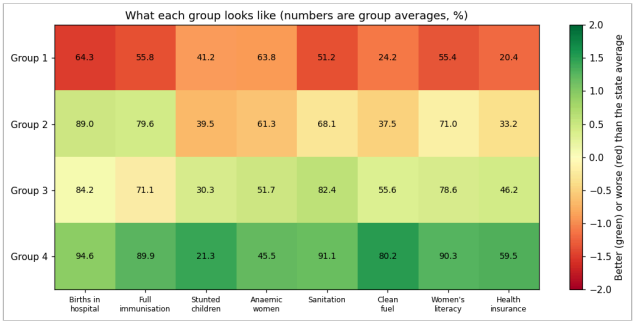


Figure 4 — What each group looks like, a heat map drawn by the program (chart4_profiles.png).

How to read it. Each row is a group; each cell shows the group's average for an indicator, coloured green when better than the state average and red when worse.

Group 1 is red almost everywhere. Group 2 has good hospital births and immunisation but poor nutrition — stunting and anaemia are orange. Group 3 is the reverse: good nutrition and sanitation, weaker immunisation. Group 4 is green throughout.

Figure 5: The groups on two indicators

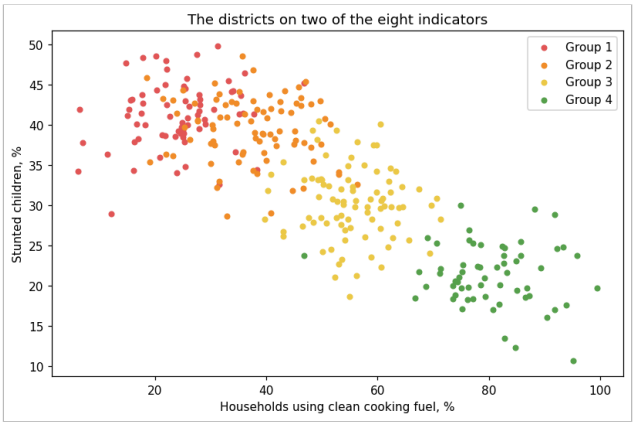


Figure 5 — The groups on two indicators, a scatter plot drawn by the program (chart5_scatter.png).

How to read it. Each dot is a district, placed by clean fuel and stunting, coloured by its group.

Even on two of the eight indicators the groups are visible, from Group 1 at top left to Group 4 at bottom right. Groups 1 and 2 overlap here — both have many stunted children — and are told apart by other indicators, such as hospital births and immunisation.

Figure 6: How well each district fits its group

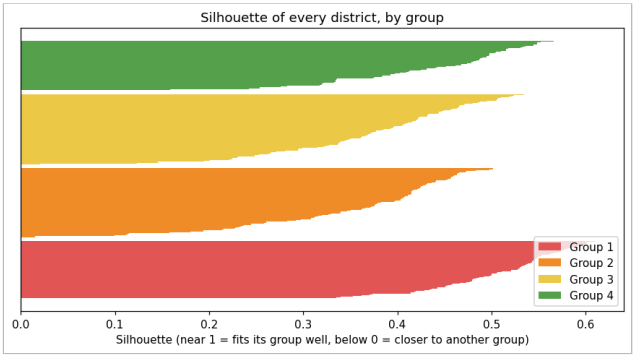


Figure 6 — How well each district fits its group, a horizontal bar chart drawn by the program (chart6_silhouette.png).

How to read it. One bar per district, grouped and sorted: the longer the bar, the more clearly the district belongs to its group.

Every district has a positive silhouette — none is closer to another group than its own. Groups 1 and 4 are the most clear-cut; Groups 2 and 3 have more districts near their borders.

DEPLOYMENT

Can you put the model into practice?

The medians, the scaler, the K-means model and the group names are saved together in district_groups.pkl. Any district can then be placed in a group by filling its blanks with the same medians, scaling it in the same way, and finding its nearest group centre.

The program tries a newly formed district whose health insurance figure is not yet known. It was placed in Group 2. Its programme focus follows from its group — stunted children, anaemic women.

In use, the state would get one page per group: its districts, its profile from chart 4, and the programmes matched to its weakest indicators — for example, immunisation drives and support for hospital births for Group 1, nutrition and anaemia programmes for Group 2.

Figure 7: Feedback: groups in the next survey

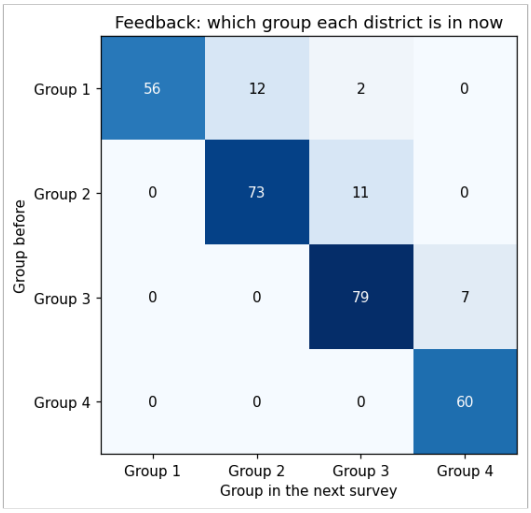


Figure 7 — Feedback: groups in the next survey, a confusion matrix drawn by the program (chart7_next_survey.png).

How to read it. Rows are each district's group before, columns its group in the next survey, using the same group centres. The diagonal is districts that stayed.

Every move is towards a better-off group, and none backwards. Most districts stayed, but Group 1 shrank, and two districts jumped from Group 1 to Group 3.

FEEDBACK

Can you get constructive feedback into answering the question?

When the next survey round arrived, each district was placed using the same group centres, so that a move between groups means real change. 32 districts changed group; 32 of them moved to a better-off group. Group 1 had 70 districts before and 56 after.

Two districts jumped two groups, from Group 1 to Group 3. These are the districts to learn from: what was done there, and could it work in the others still in Group 1?

Keeping the old centres measures progress. But as districts improve, the groups drift out of date; at some point the state should cluster afresh on the new survey, and the planner should check that the new groups still make sense.

ETHICAL CONSIDERATIONS

Groups are not labels for people

Calling a district 'Group 1' must never become a stigma for the people who live there. The groups describe needs in order to direct help, not to judge.

Averages hide people

A district in Group 4 can still have villages, castes or tribes far behind. Clustering districts does not replace attention to the most excluded within each district.

The data decides the groups

Choosing which indicators to use is a human decision that shapes the groups. It should be made openly, with health experts, and explained.

Fair use of funds

Groups should guide which programmes go where — not reduce help to districts that are improving before the improvement is secure.

THE PROJECT IN ORANGE

Orange has K-means with a built-in silhouette table, which suits this project well. The tool is **Orange Data Mining**.

WIDGET	WHAT IT DOES
File	Load districts.csv; mark district as a meta attribute (not a feature).
Impute	Fill blanks with the median (after fixing the values above 100 in a spreadsheet).
Preprocess	Normalise features to mean 0 and standard deviation 1.
k-Means	Try k from 2 to 8; Orange shows the silhouette score for each and selects the best.
Silhouette Plot	See how well each district fits its cluster.
Box Plot / Data Table	Group by cluster to describe each group's profile.

TECHNICAL DETAILS

Programming language	Python 3
Libraries	pandas, NumPy, Matplotlib, scikit-learn, joblib
Where to run it	Jupyter Notebook or Google Colab (district_health.ipynb), or any Python IDE (district_health.py)
No-code alternative	Orange Data Mining
Model type	Learning based → Unsupervised learning → Clustering
Charts produced	7
Test cases run	10

Setting it up

1. Install Python 3 (Anaconda includes Jupyter and most libraries), or open Google Colab in a browser.
2. Install the libraries: **pip install pandas numpy matplotlib scikit-learn**
3. Keep the program and its data files together in one folder.
4. Run **python district_health.py**, or open **district_health.ipynb** and run each cell in order.

The library calls it uses

CALL	STEP	WHAT IT DOES
`drop_duplicates(subset="district")`	Step 6	Removes the district entered twice
`data.loc[data[col] > 100, col] / 10`	Step 6	Fixes a missing decimal point
`fillna(median())`	Step 6	Fills blank cells
`StandardScaler().fit_transform()`	Step 6	Puts every indicator on the same scale
`KMeans(n_clusters=k, n_init=10)`	Step 7	The clustering, from ten starting points
`km.inertia_`	Step 7	For the elbow method
`silhouette_score() / silhouette_samples()`	Steps 7–8	How clearly separated the groups are
`adjusted_rand_score()`	Step 8	Checks that the groups do not change between runs
`groupby("group").mean()`	Step 8	Each group's profile
`model.predict()`	Steps 9–10	Places a new or updated district in a group

SOURCE CODE

The complete program, laid out as the ten steps of the methodology. Every figure and number in this documentation came from running it.

`district_health.py`

```
# ------------------------------------------------------------------------
# district_health.py
#
# Capstone project: grouping districts with similar health needs, so that
# each group gets the programme it needs instead of one plan for all.
#
# CBSE Class XII, Artificial Intelligence (843), Part C - Capstone Project.
# SDG 3 (Good Health and Well-being) and SDG 10 (Reduced Inequalities).
#
# Data Science Methodology, 10 steps. UNSUPERVISED LEARNING: the districts
# have no labels. K-means clustering finds the groups; the number of groups
# is chosen with the elbow method and the silhouette score, and the result
# is checked for stability. Each group is then described by its weakest
# indicators.
#
# Libraries: pandas, numpy, matplotlib, scikit-learn, joblib.
# ------------------------------------------------------------------------

# %% Step 1 - Business Understanding
```

```
print("=" * 64)
print("DISTRICTS WITH SIMILAR HEALTH NEEDS")
print("=" * 64)
print("Problem : one state-wide plan treats every district the same,")
print("          though their health needs are very different")
print("Goal    : find groups of districts with similar needs, and what")
print("          each group needs most")
print()

# %% Step 2 - Analytic Approach
print("Analytic approach : descriptive - clustering (unsupervised learning)")
print("There are no labels: nobody has said which district belongs where")
print()

# %% Step 3 - Data Requirements
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score, silhouette_samples, adjusted_rand_score
import joblib

INDICATORS = ["institutional_births_pct", "full_immunisation_pct", "stunted_children_pct", "anaemic_women_pct",
              "improved_sanitation_pct", "clean_fuel_pct", "women_literacy_pct", "health_insurance_pct"]
SHORT = ["Births in\nhospital", "Full\nimmunisation", "Stunted\nchildren", "Anaemic\nwomen",
         "Sanitation", "Clean\nfuel", "Women's\nliteracy", "Health\ninsurance"]
LOWER_IS_BETTER = ["stunted_children_pct", "anaemic_women_pct"]

# %% Step 4 - Data Collection
data = pd.read_csv("districts.csv")
print("Districts collected :", len(data))
print("Indicators          :", len(INDICATORS))
print()

# %% Step 5 - Data Understanding
print("--- Step 5: Data Understanding ---")
print("Blank cells in each indicator:")
print(data[INDICATORS].isnull().sum().to_string())
print("Values above 100 percent :", int((data[INDICATORS] > 100).sum().sum()))
print()

# Chart 1: how the indicators move together.
corr = data[INDICATORS].corr()
plt.figure(figsize=(8, 6.5))
plt.imshow(corr, cmap="RdBu", vmin=-1, vmax=1)
for a in range(len(INDICATORS)):
    for b in range(len(INDICATORS)):
        plt.text(b, a, round(corr.iloc[a, b], 2), ha="center", va="center", fontsize=8)
plt.xticks(range(len(INDICATORS)), SHORT, fontsize=8)
plt.yticks(range(len(INDICATORS)), SHORT, fontsize=8)
plt.colorbar(label="Correlation")
plt.title("How the indicators move together")
plt.tight_layout()
plt.savefig("chart1_correlation.png")
plt.close()

# %% Step 6 - Data Preparation
print("--- Step 6: Data Preparation ---")
before = len(data)
data = data.drop_duplicates(subset="district").reset_index(drop=True)
print("District entered twice, removed :", before - len(data))
typo = data[INDICATORS] > 100
print("Values typed without the decimal point, fixed :", int(typo.sum().sum()))
for col in INDICATORS:
    data.loc[data[col] > 100, col] = data[col] / 10        # 912 -> 91.2
medians = data[INDICATORS].median()
print("Blank cells filled with the indicator's median :", int(data[INDICATORS].isnull().sum().sum()))
data[INDICATORS] = data[INDICATORS].fillna(medians)
print()
print("Spread (standard deviation) of each indicator:")
print(data[INDICATORS].std().round(1).to_string())
# K-means measures distance. Without scaling, the indicators that vary most
# (clean fuel, insurance) would decide the groups on their own.
scaler = StandardScaler()
X = scaler.fit_transform(data[INDICATORS])
print("Every indicator rescaled to mean 0 and standard deviation 1")
print()

# Chart 2: the spread of every indicator.
plt.figure(figsize=(10, 4.5))
plt.boxplot([data[c] for c in INDICATORS])
plt.xticks(range(1, len(INDICATORS) + 1), SHORT, fontsize=8)
plt.ylabel("Percent")
plt.title("The eight indicators across all districts")
plt.tight_layout()
plt.savefig("chart2_spread.png")
plt.close()

# %% Step 7 - Modelling
print("--- Step 7: Modelling ---")
```

```
print("k   inertia  silhouette")
inertias = []
silhouettes = []
ks = range(2, 9)
for k in ks:
    km = KMeans(n_clusters=k, n_init=10, random_state=0).fit(X)
    inertias.append(km.inertia_)
    silhouettes.append(silhouette_score(X, km.labels_))
    print(k, " ", round(km.inertia_, 1), "  ", round(silhouettes[-1], 4))
# When two values of k have almost the same silhouette, the larger one is
# kept: more groups mean more targeted programmes, at no cost in fit.
best = max(silhouettes)
K = max(k for k, s in zip(ks, silhouettes) if s >= best - 0.01)
print("Highest silhouette :", round(best, 4))
print("Number of groups chosen :", K)
print("(the most groups with a silhouette within 0.01 of the highest)")
model = KMeans(n_clusters=K, n_init=10, random_state=0).fit(X)
print()

# Chart 3: the elbow and the silhouette.
plt.figure(figsize=(10, 4))
plt.subplot(1, 2, 1)
plt.plot(list(ks), inertias, marker="o", color="#4e79a7")
plt.axvline(K, color="gray", linestyle="--")
plt.title("Elbow method: inertia")
plt.xlabel("Number of groups (k)")
plt.subplot(1, 2, 2)
plt.plot(list(ks), silhouettes, marker="o", color="#e15759")
plt.axvline(K, color="gray", linestyle="--")
plt.title("Silhouette score")
plt.xlabel("Number of groups (k)")
plt.tight_layout()
plt.savefig("chart3_elbow_silhouette.png")
plt.close()

# Name the groups in order of need: Group 1 has the greatest need.
sign = np.array([-1 if c in LOWER_IS_BETTER else 1 for c in INDICATORS])
wellbeing = (model.cluster_centers_ * sign).mean(axis=1)    # high = better off
order = np.argsort(wellbeing)
group_of = {old: new + 1 for new, old in enumerate(order)}
data["group"] = [group_of[g] for g in model.labels_]

# %% Step 8 - Evaluation
print("--- Step 8: Evaluation ---")
print("Silhouette score :", round(silhouette_score(X, model.labels_), 3))
same = []
for seed in range(1, 11):
    other = KMeans(n_clusters=K, n_init=10, random_state=seed).fit(X)
    same.append(adjusted_rand_score(model.labels_, other.labels_))
print("Lowest agreement with 10 runs from other starting points (adjusted Rand index) :", round(min(same), 3))
print("Districts in each group:")
print(data["group"].value_counts().sort_index().to_string())
print("Group sizes :", ", ".join(str(n) for n in data["group"].value_counts().sort_index()))
print()
profile = data.groupby("group")[INDICATORS].mean().round(1)
print("Average of each indicator in each group:")
print(profile.T.to_string())
print()
state = data[INDICATORS].mean()
need = {}
for g in profile.index:
    gap = (profile.loc[g] - state) / data[INDICATORS].std() * sign   # negative = worse than average
    weakest = gap.sort_values().index[:2]
    need[g] = [w for w in weakest if gap[w] < 0]
    names = [SHORT[INDICATORS.index(w)].replace("\n", " ").lower() for w in need[g]]
    print("Group", g, "- weakest compared with the state :", ", ".join(names) if names else "none - above average on all")
sil = silhouette_samples(X, model.labels_)
data["silhouette"] = sil
print("Districts that fit their group poorly (silhouette below 0) :", int((sil < 0).sum()))
print()

# Chart 4: each group's profile.
z = (profile - state) / data[INDICATORS].std() * sign
plt.figure(figsize=(10, 3.2 + 0.4 * K))
plt.imshow(z.values, cmap="RdYlGn", vmin=-2, vmax=2, aspect="auto")
for a in range(K):
    for b in range(len(INDICATORS)):
        plt.text(b, a, profile.iloc[a, b], ha="center", va="center", fontsize=9)
plt.xticks(range(len(INDICATORS)), SHORT, fontsize=8)
plt.yticks(range(K), ["Group " + str(g) for g in profile.index])
plt.colorbar(label="Better (green) or worse (red) than the state average")
plt.title("What each group looks like (numbers are group averages, %)")
plt.tight_layout()
plt.savefig("chart4_profiles.png")
plt.close()

# Chart 5: the groups on two of the indicators.
colours = ["#e15759", "#f28e2b", "#edc948", "#59a14f", "#4e79a7", "#b07aa1", "#9c755f", "#bab0ac"]
plt.figure(figsize=(7.5, 5))
for g in range(1, K + 1):
    part = data[data["group"] == g]
```

```
    plt.scatter(part["clean_fuel_pct"], part["stunted_children_pct"], s=18, color=colours[g - 1], label="Group " + str(g))
plt.xlabel("Households using clean cooking fuel, %")
plt.ylabel("Stunted children, %")
plt.title("The districts on two of the eight indicators")
plt.legend()
plt.tight_layout()
plt.savefig("chart5_scatter.png")
plt.close()

# Chart 6: how well each district fits its group.
plt.figure(figsize=(8, 4.5))
y0 = 0
for g in range(1, K + 1):
    s = np.sort(data.loc[data["group"] == g, "silhouette"].values)
    plt.barh(range(y0, y0 + len(s)), s, height=1.0, color=colours[g - 1], label="Group " + str(g))
    y0 = y0 + len(s) + 5
plt.axvline(0, color="black", linewidth=0.8)
plt.title("Silhouette of every district, by group")
plt.xlabel("Silhouette (near 1 = fits its group well, below 0 = closer to another group)")
plt.yticks([])
plt.legend(loc="lower right")
plt.tight_layout()
plt.savefig("chart6_silhouette.png")
plt.close()

# %% Step 9 - Deployment
print("--- Step 9: Deployment ---")
joblib.dump({"medians": medians, "scaler": scaler, "model": model, "group_of": group_of}, "district_groups.pkl")
saved = joblib.load("district_groups.pkl")


def group_for(values):
    row = pd.DataFrame([values], columns=INDICATORS).fillna(saved["medians"])
    label = saved["model"].predict(saved["scaler"].transform(row))[0]
    return saved["group_of"][label]


new_district = [78, 64, 36, 60, 64, 33, 66, None]
g = group_for(new_district)
print("A newly formed district, insurance figure not yet known : Group", g)
print("Programme focus for Group", g, ":", ", ".join(SHORT[INDICATORS.index(w)].replace("\n", " ").lower() for w in need[g]))
print()

# %% Step 10 - Feedback
print("--- Step 10: Feedback ---")
nxt = pd.read_csv("districts_next_survey.csv")
nxt["group"] = [group_for(list(r)) for r in nxt[INDICATORS].values]
both = data[["district", "group"]].merge(nxt[["district", "group"]], on="district", suffixes=("_before", "_after"))
moved = both[both["group_before"] != both["group_after"]]
print("Districts in the next survey        :", len(nxt))
print("Districts that changed group        :", len(moved))
print("  moved to a better-off group       :", int((moved["group_after"] > moved["group_before"]).sum()))
print("  moved to a group in greater need  :", int((moved["group_after"] < moved["group_before"]).sum()))
print("Districts in Group 1 before         :", int((both["group_before"] == 1).sum()))
print("Districts in Group 1 after          :", int((both["group_after"] == 1).sum()))
print("Districts that jumped two or more groups:")
jump = moved[moved["group_after"] - moved["group_before"] >= 2]
for _, r in jump.iterrows():
    print("  ", r["district"], ": Group", r["group_before"], "-> Group", r["group_after"])

# Chart 7: where each district went.
table = pd.crosstab(both["group_before"], both["group_after"]).reindex(index=range(1, K + 1), columns=range(1, K + 1), fill_value=0)
plt.figure(figsize=(6, 5))
plt.imshow(table.values, cmap="Blues")
for a in range(K):
    for b in range(K):
        plt.text(b, a, table.iloc[a, b], ha="center", va="center",
                 color="white" if table.iloc[a, b] > table.values.max() / 2 else "black")
plt.xticks(range(K), ["Group " + str(g) for g in range(1, K + 1)])
plt.yticks(range(K), ["Group " + str(g) for g in range(1, K + 1)])
plt.xlabel("Group in the next survey")
plt.ylabel("Group before")
plt.title("Feedback: which group each district is in now")
plt.tight_layout()
plt.savefig("chart7_next_survey.png")
plt.close()
```

OUTPUT OF A REAL RUN

Output — python district_health.py

```
============================================================
DISTRICTS WITH SIMILAR HEALTH NEEDS
============================================================
Problem : one state-wide plan treats every district the same,
          though their health needs are very different
Goal    : find groups of districts with similar needs, and what
          each group needs most

Analytic approach : descriptive - clustering (unsupervised learning)
```

```
There are no labels: nobody has said which district belongs where

Districts collected : 301
Indicators          : 8

--- Step 5: Data Understanding ---
Blank cells in each indicator:
institutional_births_pct    0
full_immunisation_pct       4
stunted_children_pct        0
anaemic_women_pct           0
improved_sanitation_pct     0
clean_fuel_pct              4
women_literacy_pct          0
health_insurance_pct        5
Values above 100 percent : 3

--- Step 6: Data Preparation ---
District entered twice, removed : 1
Values typed without the decimal point, fixed : 3
Blank cells filled with the indicator's median : 12

Spread (standard deviation) of each indicator:
institutional_births_pct    12.4
full_immunisation_pct       13.1
stunted_children_pct         8.6
anaemic_women_pct            8.7
improved_sanitation_pct     15.9
clean_fuel_pct              21.3
women_literacy_pct          13.3
health_insurance_pct        15.6
Every indicator rescaled to mean 0 and standard deviation 1

--- Step 7: Modelling ---
k   inertia  silhouette
2   1197.7    0.3996
3   743.4    0.4115
4   514.9    0.4109
5   491.2    0.3383
6   476.6    0.2407
7   453.1    0.1968
8   437.8    0.1291
Highest silhouette : 0.4115
Number of groups chosen : 4
(the most groups with a silhouette within 0.01 of the highest)

--- Step 8: Evaluation ---
Silhouette score : 0.411
Lowest agreement with 10 runs from other starting points (adjusted Rand index) : 1.0
Districts in each group:
group
1    70
2    84
3    86
4    60
Group sizes : 70, 84, 86, 60

Average of each indicator in each group:
group                        1     2     3     4
institutional_births_pct  64.3  89.0  84.2  94.6
full_immunisation_pct     55.8  79.6  71.1  89.9
stunted_children_pct      41.2  39.5  30.3  21.3
anaemic_women_pct         63.8  61.3  51.7  45.5
improved_sanitation_pct   51.2  68.1  82.4  91.1
clean_fuel_pct            24.2  37.5  55.6  80.2
women_literacy_pct        55.4  71.0  78.6  90.3
health_insurance_pct      20.4  33.2  46.2  59.5

Group 1 - weakest compared with the state : births in hospital, full immunisation
Group 2 - weakest compared with the state : stunted children, anaemic women
Group 3 - weakest compared with the state : full immunisation
Group 4 - weakest compared with the state : none - above average on all
Districts that fit their group poorly (silhouette below 0) : 0

--- Step 9: Deployment ---
A newly formed district, insurance figure not yet known : Group 2
Programme focus for Group 2 : stunted children, anaemic women

--- Step 10: Feedback ---
Districts in the next survey        : 300
Districts that changed group        : 32
  moved to a better-off group       : 32
  moved to a group in greater need  : 0
Districts in Group 1 before         : 70
Districts in Group 1 after          : 56
Districts that jumped two or more groups:
   District 118 : Group 1 -> Group 3
   District 392 : Group 1 -> Group 3
```

TESTING

The program was tested by running it. Every case gives the whole program a small input of its own and records what actually came back.

TEST CASE	EXPECTED	ACTUAL RESULT	RESULT
Every district read	`Districts collected : 301`	`Districts collected : 301`	Pass
The district entered twice is removed	`District entered twice, removed : 1`	`District entered twice, removed : 1`	Pass
Percentages typed without the decimal point are fixed	`Values typed without the decimal point, fixed : 3`	`Values typed without the decimal point, fixed : 3`	Pass
Four groups: the most with a silhouette within 0.01 of the best	`Number of groups chosen : 4`	`Number of groups chosen : 4`	Pass
The groups do not depend on K-means' starting point	`Lowest agreement with 10 runs from other starting points (adjusted Rand index) : 1.0`	`Lowest agreement with 10 runs from other starting points (adjusted Rand index) : 1.0`	Pass
A new district with a missing figure is still placed in a group	`A newly formed district, insurance figure not yet known : Group 2`	`A newly formed district, insurance figure not yet known : Group 2`	Pass
Three clearly separate groups: three are found	`Number of groups chosen : 3`	`Number of groups chosen : 3`	Pass
Same figures in the next survey: no district changes group	`Districts that changed group : 0`	`Districts that changed group : 0`	Pass
One value typed as 600-something is fixed	`Values typed without the decimal point, fixed : 1`	`Values typed without the decimal point, fixed : 1`	Pass
The data file is missing	`FileNotFoundError: [Errno 2] No such file or directory: 'districts.csv'`	`FileNotFoundError: [Errno 2] No such file or directory: 'districts.csv'`	Pass

TEAM ROLES AND TIMELINE

Division of work

ROLE	RESPONSIBLE FOR
Project lead	Business understanding; interviews with planners and officers
Data lead	The survey indicators; checking and cleaning them
Analyst	Correlations, spreads and scaling
Modeller	K-means, the elbow, the silhouette and the stability check
Documentation and video	Group profiles, the next-survey feedback, the report and the video

Timeline

WHEN	METHODOLOGY STEPS	DELIVERABLE
June – July	1 Business Understanding, 2 Analytic Approach	Problem statement; interviews
August	3 Data Requirements, 4 Data Collection	Indicators chosen; district data collected
September	5 Data Understanding, 6 Data Preparation	Errors fixed; indicators scaled
October	7 Modelling, 8 Evaluation	k chosen; groups checked and described
November	9 Deployment, 10 Feedback	Group pages; next survey compared
December – January	Documentation and video	Report and 3-minute video

RESULTS AND FUTURE SCOPE

What the project found

1. The districts form 4 groups: the elbow and the silhouette agree.
2. The groups are stable: the lowest agreement across ten restarts was 1.0.
3. Two middle groups are behind for opposite reasons — nutrition in one, immunisation in the other — which a ranking would hide.
4. Group 1 is behind on almost everything, worst on births in hospital, full immunisation.
5. In the next survey, 32 districts moved to a better-off group and none moved back.

Limitations

- Sample data for imaginary districts
- Eight indicators only; others would change the groups
- K-means assumes round, similar-sized groups
- District averages hide inequality within districts

Future scope

- Use real NFHS district data for one state

- Add more indicators, chosen with health experts
- Cluster blocks within districts, where data allows
- Compare K-means with hierarchical clustering

What the group had to learn

- Unsupervised learning and clustering
- K-means, the elbow method and the silhouette score
- Why scaling matters for distance-based methods
- Checking stability with the adjusted Rand index
- Turning clusters into a description people can act on

CONCLUSION

The capstone found four groups of districts with different health needs, with no labels to learn from. The number of groups was not guessed but chosen with the elbow method and the silhouette score, and the groups came out the same from every starting point.

Clustering's value was in what the groups meant. Two groups in the middle of the state turned out to be behind for opposite reasons, and need different programmes — something a single ranking would have hidden.

The next survey showed districts moving to better-off groups, and two districts jumping ahead. Clustering does not improve health by itself; it helps the state put the right programme in the right place, and see where it is working.

VIDEO SCRIPT

Part C requires a video of exactly three minutes with six parts: the problem statement, the SDG the project is aligned to, the AI concepts and algorithms used, how the project works, the conclusion, and an acknowledgement to the teacher. The script below is timed to the second.

TIME	PART	WHAT TO SAY AND SHOW
0:00 – 0:25	a. Problem statement	Show a map of districts. "One plan for the whole state treats every district the same — but their needs are not the same."
0:25 – 0:40	b. SDG alignment	SDG 3 and SDG 10 icons. "Health for all, and no district left behind."
0:40 – 1:15	c. AI concepts and algorithms	"Unsupervised learning: K-means clustering, with the elbow method and the silhouette score to choose the number of groups."
1:15 – 2:15	d. How it works	Chart 3: four groups chosen. Chart 4: what each group needs. Then a new district placed in its group, and the next survey's moves.
2:15 – 2:45	e. Conclusion	"Four groups, four kinds of need — and a way to see which districts are moving ahead."
2:45 – 3:00	f. Acknowledgement	"We thank our AI teacher, [name], and the health officers who guided us."

VIVA PREPARATION

The viva voce is based on the Capstone Project. These are the questions the group prepared for, with the answers each member should be able to give in their own words.

Q1. Why is this unsupervised learning?
The districts have no labels — no one has said which group each belongs to. The model finds the groups itself.

Q2. How does K-means work?
It places k centres, assigns each district to the nearest centre, moves each centre to the average of its districts, and repeats until the groups stop changing.

Q3. How did you choose k?
With the elbow method (where inertia stops falling steeply) and the silhouette score (how clearly the groups are separated). Both pointed to four.

Q4. Why did you scale the data?
K-means measures distances. Without scaling, indicators that vary a lot, like clean fuel, would decide the groups on their own.

Q5. What does a silhouette score of about 0.41 mean?
The groups are reasonably well separated but not perfectly — districts form a range rather than separate islands.

Q6. Why did you run K-means from ten other starting points?
K-means can give different groups from different starts. All ten agreed with the original, so the groups are stable.

Q7. Why not just rank the districts?
A rank is one number. Two districts can be equally behind for different reasons and need different programmes.

Q8. Which SDGs does it align to?
SDG 3, Good Health and Well-being, and SDG 10, Reduced Inequalities.

GLOSSARY

The terms used in this documentation, as the Class XII handbook defines them.

TERM	MEANING
Data Science Methodology	A prescribed sequence of iterative steps (ten, in the IBM methodology) that data scientists follow to approach a problem and find a solution.
5W1H Problem Canvas	Who, What, Where, When, Why and How — the questions used in Business Understanding to define a problem fully.
Feature	An input column the model learns from.
Target / label	The column the model must predict.
Training data	The part of the data the model learns from.
Testing data	Data held back from training, used to judge the model on rows it has never seen.
Train-test split	Dividing the data once into training and testing sets, commonly 80 : 20.
K-fold cross validation	Splitting the data into k folds and training k times, each time testing on a different fold, so every row is tested once.
Overfitting	When a model learns the training data too closely, including its noise, and does worse on new data.
Data preparation	Cleaning, combining and transforming data into a form a model can use — usually the longest step.
Feature engineering	Selecting, modifying or creating features from raw data to improve a model.
Deployment	Putting the model into practice where its users can apply it.
Feedback	Results and comments from using the deployed model, used to refine it — the step that makes the methodology a cycle.
Clustering	Unsupervised learning that groups similar rows together without any labels.
Centroid	The centre of a cluster: the average of the rows in it.
Silhouette score	How well each row fits its own cluster compared with the nearest other cluster, from −1 to 1.

REFERENCES

1. National Family Health Survey (NFHS-5) district factsheets, MoHFW and IIPS — https://rchiips.org/nfhs/
2. National Health Mission — https://nhm.gov.in/
3. NITI Aayog, Aspirational Districts Programme — https://www.niti.gov.in/
4. scikit-learn: Clustering (K-means, silhouette) — https://scikit-learn.org/stable/modules/clustering.html
5. CBSE, Artificial Intelligence (Subject Code 843), Class XII curriculum, 2026-27 — Part C.
6. CBSE, Artificial Intelligence Student Handbook, Class XII — Unit 2: Data Science Methodology.
7. Rollins, J. B. (2015). Foundational Methodology for Data Science. IBM Analytics.
8. scikit-learn documentation — https://scikit-learn.org/stable/
9. pandas, NumPy and Matplotlib documentation.
10. LambdaLab — https://www.lambdalab.in

APPENDIX: THE DATA

The first rows of each data file, exactly as the program read them.

districts.csv — the first 40 rows

DISTRICT	INSTITUTIONAL_BIRTHS_PCT	FULL_IMMUNISATION_PCT	STUNTED_CHILDREN_PCT	ANAEMIC_WOMEN_PCT	IMPROVED_SANITATION_PCT	CLEA
District 101	93.5	81.1	34.4	54.9	72.7	37.7
District 102	63.9	59.4	40.1	57.0	51.5	16.8
District 103	83.9	71.5	32.0	43.4	92.9	67.4
District 104	98.8	93.1	18.4	52.1	93.8	73.5
District 105	84.6	64.4	31.9	54.0	84.6	55.6
District 106	65.3	50.3	45.8	69.0	50.2	25.8
District 107	65.6	52.2	45.2	65.2	58.2	
District 108	95.1	86.5	13.5	48.6	90.3	82.8
District 109	83.8	70.2	35.0	49.4	85.9	57.6
District 110	70.2	81.0	40.2	68.7	63.7	34.8
District 111	64.2	52.0	42.0	68.0	55.1	30.6
District 112	99.3	71.2	26.9	44.7	76.5	54.1
District 113	90.3	69.7	31.7	48.7	78.8	62.3
District 114	99.5	81.7	38.4	53.6	63.3	38.1
District 115	87.1	73.1	29.9	49.0	85.0	61.2
District 116	66.1	53.8	43.0	57.0	64.3	25.5
District 117	86.3	62.4	38.2	55.2	65.9	26.3
District 118	573	43.5	40.7	62.1	45.6	28.9
District 119	99.5	93.0	18.8	42.8	85.5	87.2
District 120	85.7	81.1	43.4	65.3	64.5	46.3
District 121	93.6	66.2	27.9	46.1	82.7	48.4
District 122	92.0	80.8	39.0	56.1	68.6	39.3
District 123	90.9	84.7	38.1	67.2	63.3	45.3
District 124	84.5	83.1	48.6	66.5	80.3	35.7
District 125	81.6	71.5	34.9	48.9	73.9	54.7
District 126	63.5	63.6	41.8	54.9	63.1	28.0
District 127	82.1	68.8	33.9	52.8	90.9	40.9
District 128	95.2	72.1	36.7	45.9	62.3	35.8
District 129	77.3	71.7	30.6	53.4	80.8	60.7
District 130	68.7	62.1	40.0	56.7	60.5	18.3
District 131	89.1	69.3	28.3	53.3	90.4	39.8
District 132	93.3	72.5	36.4	60.2	62.7	22.0
District 133	95.2	96.3	20.1	40.1	92.5	78.5
District 134	99.4	89.5	23.1	34.9	87.7	84.9
District 135	81.2	80.4	35.4	51.8	82.2	43.1
District 136	88.9	85.6	16.1	44.0	97.8	90.5

District 137	86.2	63.4	37.6	55.4	63.2	49.9
District 138	77.2	71.6	18.7	48.1	81.6	55.0
District 139	92.2	83.1	39.3	55.3	51.9	42.3
District 140	88.1	80.7	35.2	70.1	66.7	30.0

districts_next_survey.csv — the first 40 rows

DISTRICT	INSTITUTIONAL_BIRTHS_PCT	FULL_IMMUNISATION_PCT	STUNTED_CHILDREN_PCT	ANAEMIC_WOMEN_PCT	IMPROVED_SANITATION_PCT	CLEA
District 101	94.9	85.6	28.6	55.7	78.0	42.1
District 102	66.4	62.9	39.5	56.0	55.4	21.6
District 103	86.1	76.6	29.7	38.7	95.9	73.7
District 104	99.5	99.5	15.0	52.8	95.3	78.8
District 105	89.3	70.7	28.1	53.4	87.6	59.7
District 106	82.7	71.0	35.1	61.9	69.4	44.2
District 107	73.7	62.5	46.4	64.9	60.9	31.6
District 108	97.5	91.1	14.1	45.5	92.3	85.8
District 109	85.0	76.8	31.8	47.9	87.5	62.9
District 110	77.7	83.8	40.4	67.6	67.4	38.5
District 111	84.3	72.6	33.1	58.9	73.8	45.1
District 112	99.5	76.6	26.1	38.8	76.9	56.5
District 113	92.5	74.2	27.6	44.6	78.9	67.2
District 114	99.5	86.1	36.3	52.6	68.3	45.4
District 115	90.3	75.9	27.8	50.3	86.8	67.7
District 116	71.3	56.9	39.6	57.1	69.7	33.3
District 117	88.7	64.3	32.1	51.0	71.5	28.4
District 118	76.0	69.1	29.7	47.2	66.6	43.1
District 119	99.5	99.5	19.9	41.9	89.4	92.6
District 120	88.5	83.4	43.1	60.4	67.6	46.8
District 121	97.8	76.0	27.8	42.2	92.0	51.6
District 122	93.2	87.6	37.7	55.2	71.2	46.9
District 123	96.8	87.8	36.2	69.0	69.6	51.6
District 124	89.0	89.3	48.1	64.1	76.8	41.8
District 125	86.2	76.0	33.3	43.9	75.0	60.4
District 126	66.8	68.4	39.9	53.2	67.6	34.9
District 127	86.4	73.0	30.9	51.5	96.6	43.8
District 128	96.9	74.9	31.6	41.7	63.2	39.6
District 129	81.0	75.3	27.1	50.0	82.5	65.3
District 130	74.8	69.1	42.2	52.3	65.8	24.8
District 131	97.2	75.3	27.6	49.1	92.2	46.2
District 132	96.1	77.2	36.3	58.0	66.2	27.0
District 133	99.5	99.5	19.7	35.2	96.7	80.0
District 134	99.5	92.1	16.9	30.4	92.0	90.0
District 135	83.5	85.3	33.1	48.8	85.7	50.0
District 136	94.9	89.7	14.6	39.2	99.5	97.5

District 137	90.1	66.1	36.2	51.6	68.2	53.5
District 138	80.7	81.5	18.3	46.6	85.7	57.9
District 139	95.3	90.5	37.1	49.8	54.6	45.2
District 140	89.2	84.5	33.6	66.6	68.9	35.2