# ✅ ✅ ✅ 🚑 MASTER PROJECT: DISEASE SURVEILLANCE PREDICTION & PREDICTIVE VACCINATION STRATEGY

## 🎯 1️⃣ Project Summary

**Objective:**  
Design an advanced *district-level disease outbreak & vaccination surveillance system* for the Government of India — delivering:

* Real-time outbreak trends
* Vaccination gap insights
* Advanced statistical tests
* ML-powered forecasting & risk flagging
* An interactive Power BI dashboard for policymakers

## ✅ 2️⃣ Real Data Sources

**A. IDSP Weekly Outbreak Data** — [IDSP Portal](https://idsp.nic.in/index4.php?lang=1&level=0&linkid=406&lid=3689)  
**B. Vaccination Coverage** — [MoHFW Bulletins](https://www.mohfw.gov.in/), NFHS-5  
**C. Population Data** — Census 2011 or [data.gov.in](https://data.gov.in/)  
**D. Weather Data (Optional)** — [IMD Portal](https://mausam.imd.gov.in/)

## ✅ 3️⃣ Final Raw Files

| **File** | **What’s in it** |
| --- | --- |
| idsp\_cases\_raw.csv | district, week, year, disease, cases\_reported |
| vaccination\_raw.csv | district, vaccination\_coverage |
| population\_data.csv | district, population |
| weather\_data.csv (optional) | district, week, rainfall\_mm, humidity |

**✅ 4️⃣ Data Folder Structure**

Integrated\_Disease\_Surveillance/

│

├── Data\_Raw/

│ ├── idsp\_cases\_raw.csv

│ ├── vaccination\_raw.csv

│ ├── population\_data.csv

│ ├── weather\_data.csv (optional)

│

├── Data\_Cleaned/

│ ├── cleaned\_cases.csv

│ ├── cleaned\_vaccination.csv

│ ├── cleaned\_population.csv

│ ├── cleaned\_weather.csv (optional)

## ✅ 5️⃣ Excel Tasks

* Import raw CSVs → clean nulls, fix district mismatches
* Derived columns:
  + incidence\_rate = (cases\_reported / population) \* 1000
  + risk\_flag = 1 if cases > threshold
* Pivot Table: Weekly/district trends, top districts
* Save final cleaned\_ CSVs for SQL

## ✅ 6️⃣ SQL Database Plan

**Schema:**

| **Table** | **Columns** |
| --- | --- |
| cases\_table | district, week, year, disease, cases\_reported |
| vaccination\_table | district, vaccination\_coverage |
| population\_table | district, population |
| weather\_table (optional) | district, week, rainfall\_mm, humidity |

**Key Queries:**

* Total cases by district/disease
* Join vaccination → low vax + high cases
* Rank districts by outbreak severity
* 4-week rolling avg
* Weather correlation

**✅ 7️⃣ Python EDA & Statistics Plan**

**EDA:**

* Weekly trend lines, heatmaps, scatter plots
* Rainfall vs vector-borne diseases (optional)

**Statistics:**

* T-Test: High vs low vax districts
* Chi-Square: Outbreak dependency
* Correlation: Weather vs diseases
* ANOVA: Compare means

**ML:**

* ARIMA/Prophet forecast
* Logistic Regression for outbreak risk
* Metrics: Accuracy, Confusion, ROC AUC

## ✅ 8️⃣ Power BI Dashboard

* KPIs: total cases, avg vaccination, affected districts
* Line chart: trends by disease
* Heatmap: outbreaks by district
* Scatter: vax % vs cases
* Risk table: high-risk flagged districts
* Filters: State, Disease, Date Range

## ✅ 9️⃣ Final Report & PPT

* Executive Summary
* Problem Statement
* Data Sources & Cleaning
* SQL & EDA Highlights
* Statistical Tests
* ML approach & metrics
* Power BI insights
* Recommendations
* Next Steps

## ✅ 🔚 Final Deliverables

| **File** | **Includes** |
| --- | --- |
| .xlsx | Raw + cleaned pivots |
| .sql | Schema + queries |
| .ipynb | EDA, stats, ML |
| .pbix | Power BI dashboard |
| .docx | Report |
| .pptx | Slide deck |