

COVID-19 Data Analysis

End-to-end analysis: Python, SQL, Power BI, Machine Learning





Project Scope

829M 388M 43M

Total Cases

Confirmed COVID-19
cases analyzed

Recoveries

Patients recovered
worldwide

Deaths

Global fatalities
tracked

187

Countries

Nations covered in
analysis

Comprehensive dataset spanning 6 WHO regions over 188 days

Three-Pillar Approach



SQL Analytics

Data cleaning, transformation
& deep analytics



Power BI

Interactive dashboards &
visual insights



Machine Learning

Predictive modeling with Python





Technology Stack

Data Infrastructure

- MySQL database
- SQL for analysis
- Power BI visualization

Machine Learning

- Python, Pandas, NumPy
- Scikit-learn
- Matplotlib
- Linear Regression ($R = 0.54$)

SQL Data Cleaning Process

01

Remove Duplicates

Eliminated redundant records

03

Handle NULL Values

Addressed missing data points

05

Join Datasets

Merged WHO region, country, case data

02

Standardize Dates

Unified date formats across dataset

04

Fill Missing Data

Completed unemployment & region info

06

Create Aggregations

Built daily, weekly, monthly views

Dashboard Insights

Confirmed Cases

Track global infection trends

Deaths Analysis

Regional mortality comparisons

Recovery Rates

Monitor healing progress

Active Cases

Current infection status

Machine Learning Prediction

Model: Linear Regression

Target: New Cases

Features:

- Date
- Confirmed cases
- Recovered
- Deaths
- Region (encoded)
- Country (encoded)



Model Performance

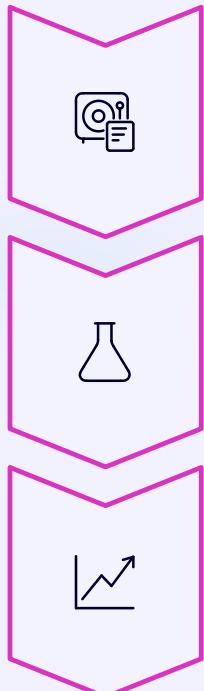


R-Value

Medium correlation achieved

COVID data exhibits high randomness – perfect prediction impossible. Model captures basic upward/downward trends despite global variability.

Getting Started



SQL Setup

Import dataset → Run cleaning queries → Execute analysis

Python Environment

Install dependencies: pip install pandas numpy scikit-learn matplotlib

Power BI Dashboard

Open .pbix file → Refresh data → Interact with visuals



Project Impact

Comprehensive Analysis

Complete pipeline from raw data to actionable insights

Visual Storytelling

Interactive dashboards reveal global pandemic patterns

Predictive Power

ML model identifies trends despite data complexity

End-to-end solution demonstrating data science expertise across multiple platforms