

Weekly Progress Summary

Shreyansh Rathaur (Team-70)

June 2025

Week 1: Back-End Data Preparation and Analysis

1. Project Kick-off and Data Ingestion

- Held the initial project meeting with Chehak and Ayza to define data sources and objectives.
- Collected raw COVID-19 datasets covering case counts, death counts, testing statistics, and vaccination records.
- Established a shared folder structure and naming convention for dataset files to ensure reproducibility.

2. Data Cleaning and Standardization

- Removed duplicate records by applying key-based de-duplication on **Country+Date** composite keys.
- Handled missing or null values using business rules:
 - Imputed missing **New Cases** by averaging adjacent days.
 - Flagged any negative or anomalous entries for manual review with Chehak.
- Standardized date formats to ISO 8601 (YYYY-MM-DD) and unified country names using a lookup table.

3. Exploratory Data Analysis

- Generated summary statistics for key fields: total cases, total deaths, tests conducted.
- Plotted time-series trends for **New Cases**, **New Deaths**, and **New Recovered** to identify spikes and anomalies.
- Calculated demographic breakdowns with initial measures:
 - **Age18to44Vaccinated**, **Age45to60Vaccinated**, **Age60PlusVaccinated**
 - **FemaleVaccinated**, **MaleVaccinated**, **TransgenderVaccinated**
- Shared EDA findings in a two-page report PDF for mentor feedback.

4. Measure Definition in Power BI

- Imported the cleaned dataset into Power BI Desktop.
- Defined and tested core measures in DAX:
 - **New Cases**, **New Deaths**, **New Recovered**
 - **Population**, **LatestUpdateDate**
 - **Global_TotalCases**, **GlobalTotalDeaths**, **Global_DeathRatePct**, **GlobalAvgTestPerPopulationPct**
 - **India_TotalCases**, **India_TotalDeaths**, **India_DeathRatePct**, **India_SpreadRatePct**
 - Operational trackers: **TotalDosesAdministered**, **TotalIndividualsVaccinated**
- Validated each measure's logic by cross-referencing against raw Excel pivot tables.

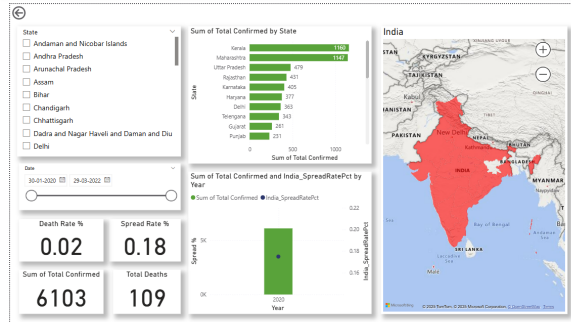
5. Iterative Testing and Validation

- Ran multiple data refresh cycles, verifying that **LatestUpdateDate** updated correctly.
- Performed boundary tests for date filters and confirmed that **New Cases** did not produce negative values.

- Coordinated a mid-week review session with the mentor to refine DAX formulas and address performance bottlenecks.

6. Documentation and Knowledge Sharing

- Created a step-by-step `README.md` documenting data ingestion, cleaning rules, and DAX definitions.
- Held a knowledge-transfer call with Chehak and Ayza to walk through the Power BI model and EDA insights.



(a) Sample Dashboard Card Layout

| From: table (column) ↑ | Relationship | To: table (column) | Status |
|-----------------------------------|--------------|----------------------------------|--------|
| country_vaccinations (country) | 1:1 | worldometer_coronavirus_sum... | Active |
| country_vaccinations (date) | 1:1 | DataTable (Date) | Active |
| country_vaccinations_by_man... | 1:1 | DataTable (Date) | Active |
| COVID-19 Global Statistics Dat... | 1:1 | worldometer_coronavirus_sum... | Active |
| COVID-19 Global Statistics Dat... | 1:1 | country_vaccinations (country) | Active |
| COVID-19 India Statewise Vacc... | 1:1 | IndiaCovidVaccination2023 (St... | Active |
| Covid_19_Countrywise_timeser... | 1:1 | DataTable (Date) | Active |
| covid_19_india (Date) | 1:1 | covid_vaccine_statewise (Upda... | Active |
| covid_19_india (Date) | 1:1 | DataTable (Date) | Active |
| covid_19_india (State/UnionTer... | 1:1 | StateBridge (State) | Active |
| covid_vaccine_statewise (State) | 1:1 | StateBridge (State) | Active |

(b) Manage Relationships & Measures

Figure 1: (a) Dashboard mockup from Week 1 measures vs. (b) Relationship configuration in Week 2.

Week 2: Data Modeling, Relationship Management, and Dashboard Preparation

1. Model Design and Table Relationships

- Reviewed data schema; identified fact tables (`CleanedData`, `VaccinationStats`) and dimension tables (`Date`, `Country`, `Demographics`).
- Configured one-to-many relationships in Power BI:
 - `Date[DateKey] → CleanedData[DateKey]`
 - `Country[CountryName] → CleanedData[Country]`
 - `Demographics[DemographicID] → VaccinationStats[DemographicID]`
- Enforced referential integrity; set cross-filter direction to single-direction; eliminated ambiguous paths.

2. External Measures and Helper Tables

- Created a `RollingWindow` helper table for 7-day moving averages.
- Defined DAX trend measures: `7DayAvgCases` and `CumulativeVaccinationPct`.
- Populated an `ExternalMeasures` table with measure metadata (name, formula, description).

3. Cross-Validation of Measures

- Performed cross-table aggregation tests to ensure consistency of `Global.TotalCases` and `India.TotalCases`.
- Verified demographic segmentation measures aggregated correctly across relationships.
- Resolved circular dependencies by adjusting filter directions.

4. Dashboard Wireframes and Prototypes

- Drafted wireframes for Global Overview, India Analysis, and Demographics pages.
- Built prototype visuals: KPI cards, filled maps, bar combo charts; refined styling per mentor feedback.

5. Performance Tuning and Optimization

- Analyzed VertiPaq storage to identify high-cardinality columns.
- Pruned unused fields; added summary tables for global aggregates.
- Achieved sub-3-second refresh on key report slices.

6. Final Validation and Handoff

- Conducted full model refresh; confirmed `LatestUpdateDate` updated correctly.
- Generated a one-page data dictionary of all measures.
- Shared the completed report with the front-end team for visual polish.

enumitem caption

Week 3: Enhanced Forecasting Page with Python Model

1. Mentor Feedback & Page Redesign

- Noted anomalies in the Actual Spread Rate chart (e.g. April–July 2020 spike, October 2020 downturn) and annotated them.
- Enhanced the narrative panel to explain these key outliers and their context.
- Swapped Power BI's built-in forecast for a Python-driven Holt–Winters model to extend the horizon and explicitly control uncertainty.

2. Final Forecasting Page Screenshot

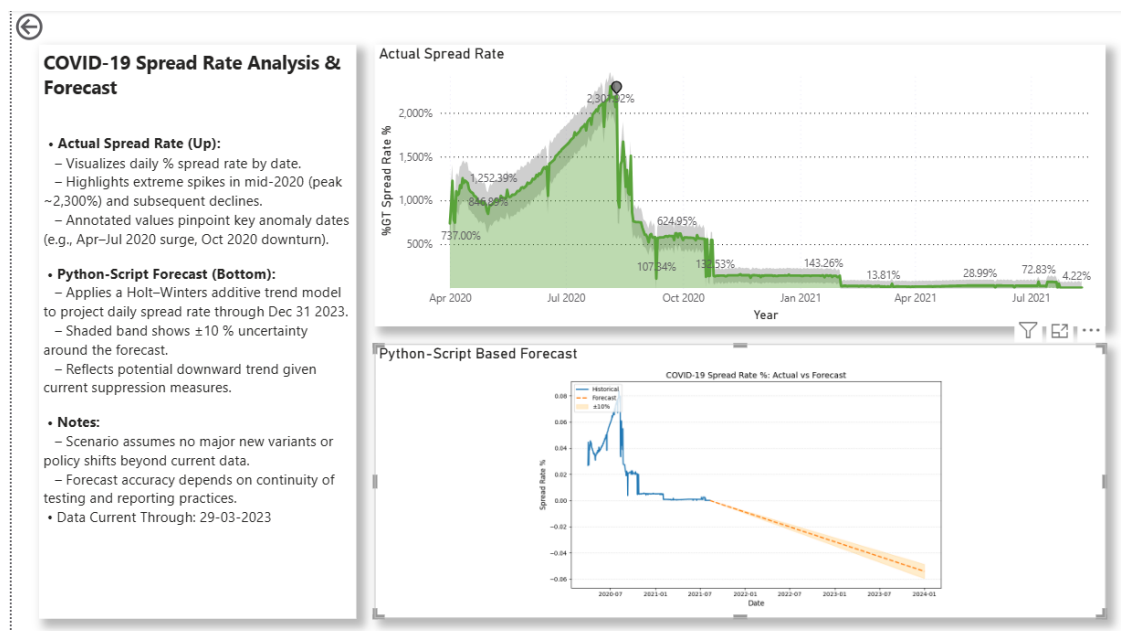


Figure 2: Power BI Forecasting page: anomaly-annotated Actual Spread Rate (top), narrative panel (left), and Python-script forecast (bottom).

3. Python-Based Holt–Winters Model

- Applied additive Holt–Winters exponential smoothing to forecast daily Spread Rate % through 31 Dec 2023.
- Overlaid a $\pm 10\%$ uncertainty band to communicate forecast confidence.
- Chose an additive trend to best capture the gradual epidemic decline under current control measures.

4. Python Script Used

```
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.holtwinters import ExponentialSmoothing

df = dataset.copy()
df['Date'] = pd.to_datetime(
    df['Month'].astype(str) + ' ' +
    df['Day'].astype(str) + ' ' +
    df['Year'].astype(str),
    format='%B %d %Y'
)
df = df.sort_values('Date')
series = df.set_index('Date')['Spread Rate %']

fit = ExponentialSmoothing(
    series, trend='add', seasonal=None,
    initialization_method='estimated'
).fit(optimized=True)

last = series.index.max()
future = pd.date_range(start=last + pd.Timedelta(days=1),
                       end='2023-12-31', freq='D')
forecast = fit.forecast(len(future))
forecast.index = future

plt.figure(figsize=(10,6))
plt.plot(series, label='Historical', linewidth=2)
plt.plot(forecast, label='Forecast', linestyle='--', linewidth=2)
plt.fill_between(
    forecast.index,
    forecast * 0.9,
    forecast * 1.1,
    color='orange', alpha=0.2, label='±10%'
)
plt.title('COVID-19 Spread Rate %: Actual vs Forecast')
plt.xlabel('Date')
plt.ylabel('Spread Rate %')
plt.legend(loc='upper left')
plt.grid(axis='y', linestyle=':', alpha=0.6)
plt.tight_layout()
plt.show()
```