THINK

MAKING A DIFFERENCE

Unlocking the Power of Data to Improve Health with AI

Prompt Sheets

District Level DS-TB Data Prompt Sheets

The prompts and examples for these sheets are based on a DHIS2 Data Export for the TB control programme.

PROMPT SHEET 1: Data Aggregation of District Level Data

- Calculating Proportions
- Sorting Data
- Presenting Data

PROMPT SHEET 2: Data Insights into District Level Data

- Describing Data
- Finding Trends
- Interpreting Data
- Programme Performance
- Insights into Predictive Analytics

PROMPT SHEET 3: Programme Strategy with District Level Data

Patient Level DR-TB Data Prompt Sheets

The prompts and examples for these sheets are based on a EDR Web Data Export for the TB control programme.

PROMPT SHEET 4: Data Aggregation of Patient Level Data

- Calculating Proportions
- Sorting Data
- Presenting Data

PROMPT SHEET 5: Data Insights into Patient Level Data

- Data Description
- Data Trends
- Data Interpretation
- Programme Performance
- Insights into Predictive Analytics
- Advanced Data Analysis
- Key findings

PROMPT SHEET 6: Programme Strategy with Patient Level Data

Prompt Sheet 1: Data Aggregation of District Level DS-TB Data

- Calculating Proportions
- Sorting Data
- Presenting Data

The prompts and examples below are based on a DHIS2 Data Export for the TB control programme. Be sure to use the indicators and descriptions as they appear in your data set.

DATA AGGREGATION

Collecting data into a summary form to allow further analysis.

Before starting, ask ChatGPT to review the data by giving a prompt such as:

- *Describe the dataset.*

Calculating proportions

- Calculate proportions for the following ***TB indicators*** per ***District or Facility***:
  - ***TB screening coverage (proportion of all facility attendants that were screened for TB)***
  - ***Clients eligible for TB test / Symptomatic clients***
  - ***TB test using GeneXpert / Investigation***
  - ***TB bacteriologically confirmed / TB confirmed.***
  - ***Proportion of Clinically diagnosed TB clients / all people diagnosed with TB.***
  - ***TB total diagnosed.***
  - ***Treatment Coverage (Nr of clients who started TB Treatment / TB Treatment Initiation***
  - ***Initial loss to follow up.***
- Display results in a table and show both raw numbers and proportions as %.

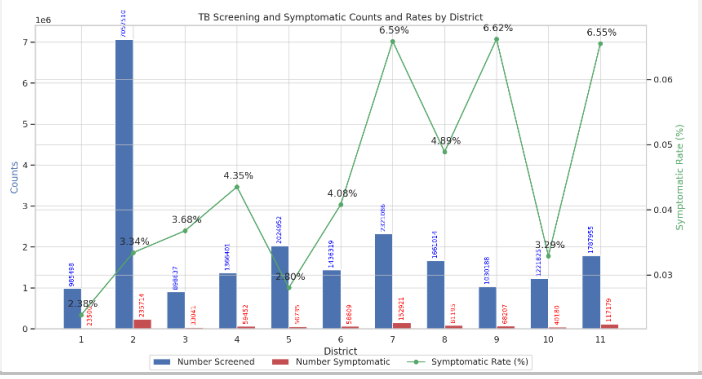
Sorting Data

- Rank the ***Districts / Facilities*** by ***Headcount / Screening rate / Number of TB investigations / TB test rate / DS-TB total diagnosed rate*** in descending order. Display the results in a table.
- Create a table of ***Headcounts, Number screened and Screening rate*** for each ***District / Facility***. Rank the ***Districts / Facilities*** in order of ***Screening Rate***.

Presenting Data

- Display this in a bar graph showing ***Number of patients screened for TB, Number symptomatic*** for each ***District / Facility***.
- Add a secondary axis for ***Symptomatic Rate*** as a line graph.

Figure 1: *Extract from Output: Display in bar graph showing Number Screened and Number Symptomatic. Adda secondary axis line graph showing the Symptomatic Rate.*



Prompt Sheet 2: Data Insights into District Level DS-TB Data

- Data Description
- Trends
- Data Interpretation
- Programme Performance
- Insights into Predictive Analytics

The prompts and examples below are based on a DHIS Data Export for DS-TB.
Be sure to use the indicators and descriptions as they appear in your data set.

DATA INSIGHTS

This is the analysis of patterns and statistics within data. Interpretation of the patterns can give insight into the TB programme and inform decision making.

Data Description

These questions may be asked after any data set upload, data aggregation, data insight, or advanced data analysis output from ChatGPT.

- Provide key findings from this data set.
- Describe the findings.
- Interpret results and provide a simple explanation.

Figure 2: *Extract from Output: Provide key finding from this data set*

ChatGPT

The results from the TB screening and treatment data can be explained simply as follows:

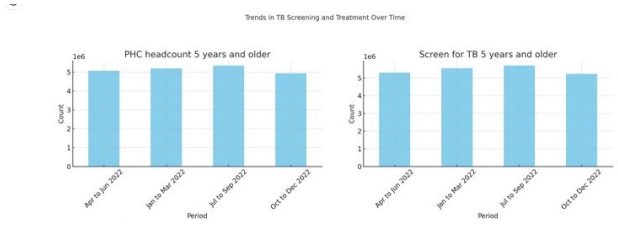
1. **High Screening Coverage**: More people were screened for TB than the total number recorded at health centers, which suggests that the TB screening program is extremely thorough.
2. **Small Fraction Show Symptoms**: Only about 4% of the people screened showed symptoms of TB, indicating that the majority of the population screened is healthy regarding TB.
3. **Efficient Follow-Up**: Nearly every symptomatic person was investigated further, showing that the health system is effectively catching and following up on potential TB cases.
4. **Low Confirmation Rate**: Out of those investigated, around 4.5% were confirmed to have TB, meaning most people who showed symptoms did not actually have TB. This indicates good selectivity in the screening process.
5. **High Treatment Initiation**: Almost all individuals confirmed with TB started treatment, showcasing strong performance in transitioning diagnosed patients into treatment programs.

Overall, the data suggests that the TB screening and treatment system is working well, with efficient screening, follow-up, and treatment initiation processes in place.

Finding Trends

- Describe the trends ***Between the Four Time Periods*** shown in the data set.
- Describe the trends in ***Headcount / Screening Numbers / Screening Rate / Number confirmed.***

Figure 3: *Extract from Output: Describe the trends in Headcount and Screening Numbers.*



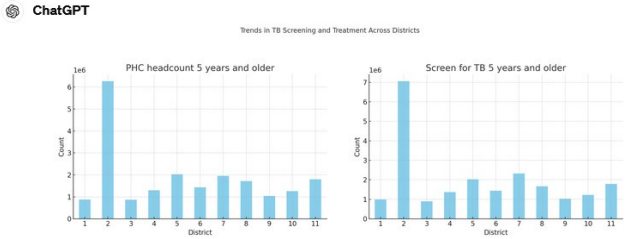
The plots display trends in TB screening and treatment over different time periods, indicating several key patterns:

1. **Primary Health Care (PHC) Headcount**: The number of individuals aged 5 years and older visiting PHC centres remains relatively consistent across the periods, showing stable engagement with health services.
2. **Screening for TB**: Screening numbers generally follow the headcount trends but are consistently higher than the headcount itself, indicating that screening protocols are robustly applied, perhaps with some individuals being screened multiple times.

- Describe how the trends vary across the ***Districts / Facilities.***

Figure 4: *Extract from Output: Describe how the trends vary across the districts.*

ChatGPT



1. **PHC Headcount (5+ years)**: The headcount in primary health care centers shows considerable variation among districts. Some districts see significantly higher headcounts, which might be due to larger populations or more frequent visits by individuals.
2. **Screening for TB (5+ years)**: Screening efforts are consistently higher across all districts compared to the headcount, with some districts exhibiting particularly high screening numbers. This could reflect differences in public health policies, the availability of screening facilities, or the emphasis placed on TB screening in certain areas.
3. **Symptomatic for TB (5+ years)**: The number of symptomatic individuals varies across districts but generally aligns proportionally with the headcount and screening numbers. Districts with more screenings tend to identify more symptomatic individuals, which is expected.
4. **TB Investigation (5+ years)**: Investigations closely follow the number of symptomatic individuals in each district, indicating a consistent approach to following up on symptomatic cases across districts.

Data Interpretation (*Arriving at conclusions from reviewing data)*

Calculate the indicators first, then proceed to prompts below.

- Calculate ***Screening Rate*** and ***Symptomatic Rate.***
- Explain what a ***Low Screening Rate / Low Symptomatic Rate / Low TB Test Rate / Low TB Diagnosed Rate / Low Treatment Initiation Rate*** indicates.
- Explain what a ***Screening Rate >100% / Eligible for Investigation / Symptomatic Rate >100% / Investigation Rate >100% / TB Confirmed Rate >100% / Treatment Initiation Rate >100%*** indicates.
- Explain the relationship between ***TB Test Rate*** and ***TB Diagnosed Rate.***

Figure 5: *Extract from Output: Explain what a Low TB Diagnosed Rate indicates*

ChatGPT

A low TB diagnosed rate, such as the 4.57% confirmed rate observed in your data, could indicate several things about the TB management system and the population being screened:

1. **Effective Screening and Investigation**: A low diagnosed rate may suggest that the initial screening and subsequent investigations are effectively distinguishing between those who likely have TB and those who do not. This helps in preventing unnecessary treatments and focuses resources on individuals who are more likely to be actually infected.
2. **Low Prevalence of TB**: It could also indicate that TB is not widely prevalent among the population being screened. This would be particularly true in contexts where public health measures and TB control strategies have been successful in reducing the incidence of the disease.
3. **High Quality of Diagnostic Techniques**: The quality and accuracy of diagnostic tests used during the investigation phase might be high, enabling precise identification of TB cases. This ensures that only true positives are confirmed, thus maintaining a low diagnosed rate.
4. **Comprehensive Healthcare Access**: Broad access to healthcare and preventive services might be reducing the incidence of TB, reflected in fewer confirmed cases among those screened and investigated.

Programme performance

- Indicate which ***District / Facility*** has the highest number of ***TB Diagnosed*** in relation to ***Headcount*** for ***the year.***
- Indicate which ***District / Facility*** has the lowest number of ***TB Diagnosed*** in relation to ***Headcount*** for ***the year.***
- Indicate which 5 ***Districts / Facilities*** have the ***Highest Positivity Yield of TB Investigations.***
- Indicate the top 5 performing ***Districts / Facilities*** based on this data set. Good performance is defined as ***Screening Rate >95%, Investigation Rate >95%***, ***Treatment Initiation Rate >95%*** and ***Excludes Screening Rate >100%*** from good performance.
- Indicate which ***Districts / Facilities*** are showing an improving trend in ***TB Programme Performance*** for ***the year.***
- Based on this data set, indicate how the TB programme performance has progressed over ***the year***.
- Based on this data set, indicate what are the strengths and weaknesses in the ***TB Programme***.

Figure 6: *Extract from Output: Indicate how the TB programme performance has progressed over the year.*

1. **Confirmed Rates**:
   - **Jan to Mar 2022**: The highest confirmed rate of 4.85%, suggesting more confirmed cases per investigation during this period.
   - **Apr to Jun 2022**: The rate decreased to 4.12%.
   - **Jul to Sep 2022**: Slightly dropped to 4.34%.
   - **Oct to Dec 2022**: Increased again to 4.78%.

   The fluctuation in confirmed rates suggests variations in either the prevalence of TB during these periods, possibly due to seasonal factors, or differences in the effectiveness of the investigative process.
2. **Treatment Start Rates**:
   - **Jan to Mar 2022**: A treatment start rate of 94.24%, which is the lowest rate for the year.
   - **Apr to Jun 2022**: Improved to 96.85%.
   - **Jul to Sep 2022**: Slightly decreased to 95.84%.
   - **Oct to Dec 2022**: Peaked at 97.11%, the highest for the year.

   This trend indicates an overall improvement in starting treatment for confirmed TB cases throughout the year, with the final quarter showing the highest rate of initiating treatment.

Insights into Predictive Analytics

*For the best results, use the **Data Insights prompts** above first.*

- Based on this data set, what could be trends in ***DS-TB confirmed numbers / DS-TB Treatment Initiation numbers*** for the ***Next Quarter / Next Year.***

Prompt Sheet 3: Programme Strategy with District Level DS-TB Data

Programme Strategy

The prompts and examples below are based on a DHIS Data Export for DS-TB.
Be sure to use the indicators and descriptions as they appear in your data set.

PROGRAMME STRATEGY

Once data has been analysed and interpreted, ChatGPT can also assist in addressing gaps and strategizing to improve TB Programme performance.

Programme Strategy Prompts

- Based on the findings, indicate which interventions may improve TB programme performance in the ***Districts***.

Figure 7: *Extract from Output: Based on the findings, indicate which interventions may improve TB programme performance in the districts*

1. **Enhanced Screening and Targeted Outreach**: Increase TB screening efforts in districts with lower screening rates, especially targeting populations that are at higher risk or less accessible. Mobile screening units and community health campaigns can be employed to reach out to underserved areas.
2. **Improved Diagnostic Technology**: Invest in more advanced diagnostic technologies, such as rapid molecular tests, to reduce the time between symptom presentation and confirmation of TB. This could help increase the confirmed rate by ensuring that symptomatic individuals are diagnosed more quickly and accurately.
3. **Data-Driven Resource Allocation**: Analyze district-specific data to tailor resource allocation better. For example, districts with lower treatment initiation rates may benefit from additional healthcare facilities or resources to support TB treatment, such as more healthcare personnel trained in TB management.
4. **Strengthen Follow-up Procedures**: Implement or enhance follow-up systems to ensure that individuals who are screened and found symptomatic are quickly investigated and, if confirmed, started on treatment without delays. This could involve integrating digital health tools that track patient progress through the TB care continuum.
5. **Community Engagement and Education**: Increase community awareness and education about TB symptoms and the importance of screening and treatment. This could involve working with local leaders and organizations to disseminate information and reduce stigma associated with TB.
6. **Cross-Sector Collaboration**: Collaborate with other sectors such as housing, education, and social services to address the social determinants of health that contribute to TB incidence. Ensuring better living conditions, nutrition, and education about health can help reduce TB risks and improve treatment outcomes.

Implementing these interventions could lead to improvements in TB programme performance by ensuring more comprehensive coverage, quicker diagnosis, and better treatment adherence, ultimately leading to lower TB incidence and better health outcomes in the community.

- Indicate how the ***Screening Rate / TB Test Rate / Treatment Start Rate / TB Programme Performance*** can be improved in ***District X / Districts / District X and Y.***
- List ways in which the ***Screening Rate >100%*** can be addressed.
- Indicate what the barriers to ***TB Programme*** performance are in the ***Districts / Province / District X.***
- List ways in which the barriers to the ***TB Programme*** can be addressed in ***District / Province / District X.***
- Indicate how the population demographics in ***District / Province*** impacts access to healthcare.
- Indicate ways ***TB Case Finding*** can be optimized in ***District / Province / District X.***

Prompt Sheet 4: Data Aggregation of Patient Level DR-TB Data

- Calculating Proportions
- Sorting Data
- Presenting Data

The prompts and examples below are based on an EDR Web Data Export.
Be sure to use the indicators and descriptions as they appear in your data set.

***PLEASE NOTE:** When using Patient level data, it is critical that Patient Identifying Information (e.g. Names) are removed and the data is anonymized.*

DATA AGGREGATION

Collecting data into a summary form to allow further analysis.

Before starting, request ChatGPT to review the data by giving a prompt such as:

- Describe the dataset.

Calculations

- Calculate proportions as a % for the following ***DR-TB indicators*** per ***District / Facility***:
  - ***HIV positive status***
  - ***Gender***
  - ***Age < 5 years***
  - ***Age > 65 years***
  - ***Patient category: New***
  - ***Patient category: previous first line treatment***
  - ***Patient category: previous second line treatment***
  - ***Pulmonary TB***
  - ***Extrapulmonary TB***
  - ***BMI under 18.5***
  - ***RR-TB***
  - ***MDR-TB***
  - ***XDR-TB***
  - ***Short Regimen***
  - ***Long / Individualised Regimen***
- Display the results in a table. Include raw numbers and proportion as %.
  - *Example: Calculate the proportion of **HIV positive** and **Extrapulmonary TB** for each **DR-TB District**. Display the results in a table and include raw numbers and proportion as %.*
- Indicate ***HIV status*** distribution by ***Age Group*** in each ***District.*** Display the results in a table including proportions as % and raw numbers.
  - Example: Create a variable called ***"Treatment Success"*** by adding ***Cured*** and ***Treatment Completed*** outcomes.
- Calculate ***Treatment Success Rate*** for each ***DR-TB District***. Rank the ***DR-TB Districts*** by ***Treatment Success Rate***. Display the results in a table including rates as % and raw numbers.
  - Example: Display the numbers and proportions of each Final Treatment Outcome. Create a variable called ***"Not Died"*** by adding the following ***final Treatment Outcomes***: ***Treatment***

Completed, ***Cured***, ***Loss to Follow Up***, ***Treatment Failure***, ***Transferred Out*** and ***Moved Out*** outcomes.

- Display ***Treatment Success Rate*** by ***HIV Status / Age Group / Gender / BMI <18.5 and BMI ≥18.5 / Regimen Type*** short or individualised ***/ Patient Category: New, Relapse, Treatment after Lost To Follow Up.*** Display this in a table. Display this in a graph.
  - Example: Display ***Treatment Success Rate*** by ***Short and Individualized Regimen***. Display this in a table and a graph.
  - Example: Display ***Treatment Success Rate*** by ***HIV status***. Display this in a table.
  - Example: Display ***Treatment Success Rate*** by ***BMI <18.5 and BMI ≥18.5***. Display this in a table.
  - Example: Display ***Lost To Follow Up Rate*** by the patient categories ***New***, ***Previous 1st line Treatment, Previous 2nd line treatment*** and ***Previous Lost To Follow Up***. Display this in a table.
- Identify the HIV positive people whose ART start date was after DR-TB treatment start date. Calculate the ***average time to ART start date.***
- Calculate average ***Time To Death*** in days.
- Calculate average ***Time To Death*** by ***HIV Status***.
- Calculate average time to death for ***BMI under 18.5 and BMI ≥18.5.***
- Calculate ***Time To Death*** by ***Age under or over 65 years***.
- Calculate the average ***Days On Treatment*** to ***Lost To Follow Up Outcome***.

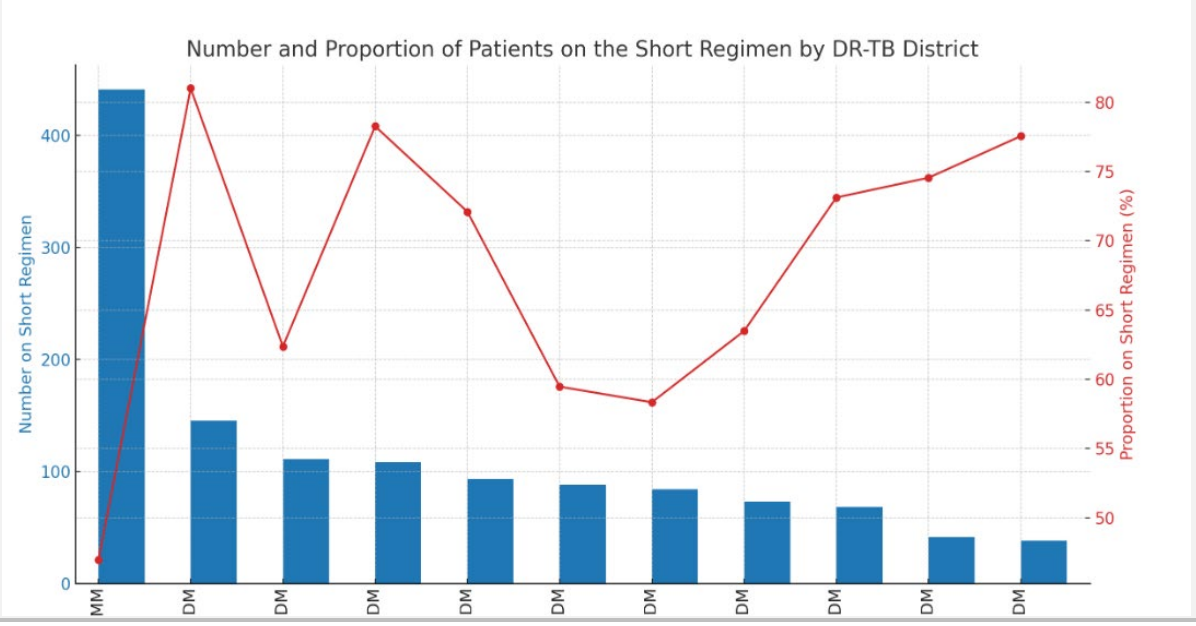
Sorting Data

- Rank the ***DR-TB Districts*** by ***cohort numbers.*** Display this in a table.
- Calculate the proportion in % on the ***Short Regimen*** for each ***DR-TB District***. Rank the districts by proportion on the ***Short Regimen*** and display this in a table.

Presenting Data

- Create a bar graph of numbers on the ***Short Regimen***, with a secondary axis for proportion on ***Short Regimen*** as a line graph.

Figure 8: *Extract from Output: Create a bar graph of numbers on the Short Regimen, with a secondary axis for proportion on Short Regimen as a line graph.*



Prompt Sheet 5: Data Insights into Patient Level DR-TB Data

- Data Description
- Data Trends
- Data Interpretation
- Programme Performance
- Insights into Predictive Analytics
- Advanced Data Analysis
- Key findings

The prompts and examples below are based on an EDR Web Data Export.
Be sure to use the indicators and descriptions as they appear in your data set.

***PLEASE NOTE:** When using Patient level data, it is critical that Patient Identifying Information (e.g. Names) are removed and the data is anonymized.*

DATA INSIGHTS

This is the analysis of patterns and statistics within data. Interpretation of the patterns can give insight into the TB programme and inform decision making.

Data Description

- Indicate the key findings from this data set.
- Interpret results and provide a simple explanation.
- *These questions may be asked after any data set upload, data aggregation, data insight, or advanced data analysis output from ChatGPT.*
  - Example: Display ***Treatment Success Rate*** *by* ***HIV Status*** *in a table. List the key findings.*

Data Trends

- Describe the trends shown in the data set.
  - Example: Indicate if there is a trend in ***Cohort Numbers*** during ***Time Period / Year.***
  - Example: Indicate if there is a seasonal pattern to ***Cohort Numbers*** or ***Treatment Outcomes.***
- Indicate the trends in ***[insert variable / indicator]***.
  - Example: Indicate the trends in ***HIV Positive Rate / Short Regimen Rates / Treatment Success Rates / LTFU Rates / Died Rates***
- Indicate how ***[insert variable / indicator]*** vary by ***[insert variable / indicator]***.
  - Example: Indicate how ***HIV Status / Age Group / Gender / BMI / Regimen Type: Short and Log (Individualized) / Final Treatment Outcomes*** vary by ***DR-TB District***.
  - Example: Indicate how ***Time in Days to LTFU Outcome*** vary by ***Age Group.***
  - Example: Indicate how final ***Treatment Outcome*** vary by ***TB Type (Pulmonary and Extrapulmonary).***
- List the ***DR-TB Districts / Facilities*** that are showing an improving trend in ***Treatment Success Rate.***

Data Interpretation (*Arriving at conclusions from reviewing data)*

- Indicate if there is an association between ***[insert variable / indicator]*** and ***[insert variable/ indicator]***.
  - Example: *Indicate if* ***BDQ-Containing Regimens*** *are associated with better* ***Treatment Success Rates***.

- Provide ***HIV distribution by gender***, include proportions and total row. Display results in a table for each variable. Indicate if there is an association between ***HIV Status*** and ***Gender***.
- Describe the relationship between ***[insert variable / indicator]*** and ***[insert variable / indicator]:***
    - HIV Status
    - Age Group
    - Gender
    - BMI
    - Regimen Choice: Short or Individualised
  - Example: Describe the relationship between culture conversion and final ***Treatment Outcomes?***
  - Example: Describe the relationship between ***BDQ-containing regimens*** and final ***Treatment Outcomes?***
  - Example: Describe the relationship between ***a Signed Consent Form*** and ***Data Completeness***.
  - Example: Describe the relationship between ***HIV Status*** and ***BMI < 18.5.***
  - Example: *Display* ***Treatment Success Rate*** *by* ***HIV Status*** *in a table.* Describe the relationship between ***Treatment Success Rate*** and ***HIV Status.***
- Indicate if ***[insert variable / indicator]*** impacts ***[insert variable / indicator]***
  - Example: Review the ***Final Treatment Outcomes*** and the ***Regimen Type: Short or Individualized.*** Indicate if the ***Regimen Type*** impacts the ***Lost to Follow Up Rates.***
  - Example: Review the ***Final Treatment Outcomes*** and ***Gender.*** Indicate if ***Gender*** impacts the ***Lost to Follow Up Rates.***
  - Example: Review ***Gender*** and ***BMI.*** Indicate if ***Gender*** impacts ***BMI.***
- Indicate if ***[insert variable / indicator]*** is influenced by ***[insert variable/ indicator].***
  - Example: Calculate ***average time to death***. Indicate if ***Time to Death*** is influenced by ***Regimen Type: Short or Individualized***
- Indicate if the difference between ***[insert variable/ indicator]*** and ***[insert variable/ indicator]*** is statistically significant.
  - Example: Indicate if the difference between ***Male*** and ***Female Gender in the Cohort*** is statistically significant.
  - Example: Indicate if the difference in ***Time to Death*** by ***HIV Status*** is statistically significant.
- Indicate if there is a statistically significant difference in ***[Insert Final Treatment Outcome]*** rates by ***HIV Status / Age Group / Gender / BMI <18.5 or ≥18.5 / Regimen type (short or individualized) / Patient Category***.

Programme performance

- List ***Districts / Facilities*** that have the best data completeness.
- Based on this data indicate how the ***DR-TB Programme Performance*** progressed between ***[insert dates or time periods]***.
- List the strengths and weaknesses in the ***DR-TB Programme*** based on this data set.
- List the gaps in ***TB Programme Performance*** based on this data.
- Describe how ***Treatment Outcomes*** compare with global benchmarks.

Insights into Predictive Analytics

*For the best results, use the **Data Insights prompts** above first followed by these prompts:*

- Indicate the predictors of a ***Treatment Success*** outcome.
- Indicate the predictors of a ***LTFU Outcome***.
- Indicate the predictors of a ***Died Outcome***.

Advanced Data Analysis

- Calculate ***odds ratio*** for ***Gender*** and ***Final Treatment Outcomes***. Show results in a table.
- Calculate and ***add odds ratios and p-values*** for each variable and display all results in a table, showing individual contingency tables.
- Create a multivariate logistic regression model using ***Death*** as an independent variable. Use the following as dependent variables; ***Previous Drug History*** and ***HIV Status.*** Create odds ratios and p-values for each variable and display all results in a table, show individual contingency tables.
- Indicate the predictors of ***Death***. Show all results in a table. Add ***RR*** and ***95% CI.***
- Provide a detailed ***STATA profile*** for all the work done.

Key Findings

*For the best results on Key Findings, use the **Data Insights prompts** above first followed by these prompts:*

- What are the key findings from this data set?
- Provide a one paragraph summary.

Prompt Sheet 6: Data Insights into Patient Level DR-TB Data

Programme Strategy

The prompts and examples below are based on an EDR Web Data Export.
Be sure to use the indicators and descriptions as they appear in your data set.

PLEASE NOTE: *When using Patient level data, it is critical that Patient Identifying Information (e.g. Names) are removed and the data is anonymized.*

PROGRAMME STRATEGY

Once data has been analysed and interpreted, ChatGPT can also assist in addressing gaps and strategizing to improve TB Programme performance.

Programme Strategy Prompts

- Based on the findings in this data set (or data aggregation / data insight / advanced data analysis), indicate which interventions may improve the DR-TB programme performance.
- Indicate how the ***Loss To Follow Up Rate / Died Rate / Extrapulmonary TB Rates / Malnourished Rates / TB Programme Performance / Adverse Event Management*** can be improved in ***District / Province***.
- List ways in which ***Missing Data Capture*** can be addressed.
- Indicate what the barriers to ***DR-TB programme performance*** are in ***District / Province.***
- List ways in which the barriers to the ***DR-TB programme*** can be addressed in ***District / Province.***
- List the population demographics in ***District / Province*** and how to they impact access to healthcare.
- Indicate ways ***DR-TB case finding*** can be optimized in ***District / Province.***
- Provide input on whether or not patient education programmes can improve ***DR-TB outcomes.***