# INSIGHTS AND QUESTIONS

# Deliverables:

The report you produce must include the following deliverables:

1. A description of the data source used.
2. Documentation of any cleaning or manipulation of data.
3. Dataset analysis, highlighting key trends and insights.
4. Justification for the allocation of funds to specific educational initiatives.
5. Data-driven marketing strategies to promote the initiatives effectively.
6. A budget plan demonstrating optimal resource utilization.
7. A summary of your analysis

# Objective:

Your primary objective is to determine how the ed-tech company can have the most impact in improving the current education situation in Punjab specifically by addressing key constraints and the funds allocated to them:

1. School level upgradation: 10 Million Rs
2. Infrastructure (toilets, boundary walls, classrooms, bldg condition): 5 Million Rs
3. Facilities such as drinking water, and playgrounds, labs: 2 Million Rs
4. More middle and high schools in rural areas with separate schools for both genders: 20 Million Rs
5. New primary level school: 10 Million Rs

# Insights:

The following tasks will be useful in helping you navigate and using the dataset as needed:

1. Determine the total number of schools present in the dataset.

**Ans) 48191 schools are in this dataset**

1. Identify the lowest students of school\_Gender(male & female) by school level of schools established in the 2000s   
   **Ans)** **Mosques have the lowest students across multiple school level in 2000s, 0 female students and 1349 male students**
2. Determine the top 5 districts with the highest number of teachers and non-teachers in schools.  
   **Ans) FAISALABAD**

**RAHIMYAR KHAN**

**RAWALPINDI**

**LAHORE**

**SARGODHA**

1. Explore the school level and medium that have the lowest number of functional classrooms.

**Ans)** **Mosques have the lowest number of functional classrooms across all school levels. With 403 in Urdu Medium, 29 in English and 81 in Hybrid Medium**

1. Create a bar chart showcasing the distribution of teachers across different schools.

**Ans) Govt Schools have the highest number of teachers, about 70.22% , followed by Model Schools with 10.12% and Community Schools with 7.75%. the remaining is diversified among Danish Adopted School, Centre of Excellence, Pilot Secondary, Junior Model Technical High, Comprehensive and M.C Local**

1. Calculate the number of male and female students (School\_Gender)

**Ans) Total students are 11,961,473. Out of which Male are 5,736,639 and female are 6,224,834**

1. Count the number of schools in rural and urban areas and calculate the percentage of schools offering Urdu and English medium education or both in each category.

**Ans) In Punjab, the majority of schools in rural areas (88.02%) choose to provide education in Urdu. On the other hand, urban areas exhibit a more diverse landscape. While 11.98% of urban schools offer both English and Urdu as mediums, a substantial 47.67% use Urdu as the primary medium, emphasizing the influence of regional languages in urban education.**

1. Identify the district with the highest enrollment in primary schools since their establishment.

**Ans)** **RAHIMYAR KHAN 257,968**

1. Determine the district with the highest enrollment in secondary schools since their establishment.

**Ans) FAISALABAD 362868**

1. Find the district with the highest enrollment in higher secondary schools since their establishment.

**Ans) FAISALABAD 99227**

1. Create a bar chart displaying the distribution of schools according to school\_level

**Ans) The bar chart displays the distribution of schools based on their school level. The majority, 66.5%, are primary schools, followed by middle schools at 17.2%. Secondary schools make up 13.8%, while a smaller proportion is categorized as higher secondary (1.6%) and mosques (0.9%).**

1. Generate a bar chart representing the number of students categorized by gender (School\_Gender). Not needed again
2. Create a table showcasing the school\_ownership with respect to schools

**Ans) Done**

1. Visualize the distribution of teachers based on their respective job posts using a bar chart.

**Ans) The bar chart visualizes the distribution of teachers based on their respective job posts. The majority, 80%, hold permanent positions, followed by those with additional charges at 13.8%. A smaller percentage has the role of "Lookafter Charge" (5.6%), while a minimal fraction falls under "N/A" (0.5%).**

1. Create a pie chart to show the percentage of vacant and filled teaching and non-teaching posts.

**Ans) "In the pie charts, we can see a significant difference between teaching and non-teaching positions. Teaching jobs are nearly all filled, with only a tiny 0.33% unfilled. On the other hand, non-teaching roles have a much higher 51.11% vacancy rate. This points to the need for more hiring and focus on non-teaching roles in education**

1. Determine the percentage of schools with satisfactory building conditions (bldg\_condition).

**Ans) 75.3% schools are in satisfactory building conditions, 1.0% of schools have buildings that are considered dangerous, 3.1% of schools have buildings that need complete repair, 3.9% of schools have partially dangerous buildings, 16.7% of schools need minor repairs.**

1. Calculate the number of schools with satisfactory security measures.

**Ans) The number of schools with satisfactory security measures is 33,871.**

1. List the names of the 5 districts with the lowest availability of drinking water facilities in schools.

**Ans) RAWALPINDI 28**

**BAHAWALNAGAR 20**

**RAHIMYAR KHAN 11**

**SARGODHA 4**

**FAISALABAD 3**

1. Identify the names of the 5 districts with the lowest availability of electricity in schools.

**Ans) Rahimyar Khan 115**

**Bahawalnagar 80**

**Rawalpindi 41**

**Sargodha 14**

**Faisalabad 4**

1. Determine the names of the/ 5 districts with the lowest presence of boundary walls in schools.

**Ans) Bahawalnagar 184**

**Rawalpindi 176**

**Faisalabad 69**

**Rahimyar Khan 40**

**Sargodha 32**

1. List the names of the 5 districts with the lowest availability of toilets in schools.

**Ans) Bahawalnagar 44**

**Rahimyar Khan17**

**Rawalpindi7**

**Sargodha 4**

**Faisalabad 2**

# Bonus Task:

To enhance your chances of promotion and showcase your skills, create a dynamic dashboard in Excel that summarizes key findings from your analysis. This will demonstrate your ability to present complex data in a user-friendly format.

Certainly, here's a prioritized budget plan based on the pointers:

**Priority 1: Primary Education Enhancement**:

* Allocate a significant portion to enhance primary education.
* Focus on infrastructure improvements, including classroom construction and facility upgrades.
* Implement technology-driven learning solutions to enhance primary education quality and accessibility.

**Priority 2: Gender-Neutral Initiatives**:

* Allocate resources to create co-educational facilities.
* Develop campaigns to encourage enrollment and retention of students, particularly in areas with gender disparities.
* Implement programs promoting gender equality in education.

**Priority 3: Secondary and Higher Secondary Education**:

* Invest in improving the quality of secondary and higher secondary education.
* Enhance infrastructure and provide additional classrooms, drinking water, and functional toilets.
* Encourage skill development programs at the higher secondary level.

**Priority 4: Remote and Underserved Areas**:

* Allocate funds to establish new schools and upgrade existing ones in remote and underserved areas.
* Implement technology-assisted learning solutions for schools in areas with limited resources.
* Address infrastructure and basic facility needs in these regions.

**Priority 5: Teacher Development**:

* Focus on teacher training and development programs.
* Invest in recruiting and retaining high-quality teaching staff, particularly in areas with teacher shortages.
* Provide incentives to attract skilled teachers to schools in need.

**Priority 6: Infrastructure Improvement**:

* Address the lowest functional classroom numbers in mosques by building more functional classrooms in these institutions.
* Prioritize the repair and maintenance of school buildings in poor condition to ensure safety.

**Priority 7: Facilities and Security**:

* Allocate funds for basic facilities like drinking water, electricity, and toilets in schools that lack them.
* Implement security measures in schools, particularly in areas where security is lacking, to ensure the safety of students and staff.

**Priority 8: Rural vs. Urban Focus**:

* Balance investments between rural and urban areas, considering the higher percentage of schools in rural areas.
* Develop targeted campaigns to promote Urdu and English medium education in both settings, aligning with local preferences.

**Priority 9: Monitoring and Evaluation**:

* Set aside a portion of the budget for monitoring and evaluation to ensure the effective use of funds and the impact of initiatives.
* Continuously assess the progress of each initiative and make adjustments as needed.

**Priority 10: Data-Driven Marketing Strategies**:

* Develop marketing strategies that raise awareness about the educational initiatives and encourage community involvement.
* Implement campaigns targeting parents, students, and community leaders to promote education and ensure the success of the initiatives.

**Plan**

1. **Step 1: Project Kick-off and Understanding**

Meet with project stakeholders and team members to understand the goals and objectives.

Clarify the dataset's scope, available resources, and timeline.

1. **Step 2: Data Gathering and Cleaning**

Collect the education dataset related to schools in Punjab, Pakistan.

Perform data cleaning and preprocessing, including handling missing data, outliers, and standardizing data formats.

1. **Step 3: Exploratory Data Analysis (EDA)**

Explore the dataset to gain insights into the current education landscape.

Conduct the tasks listed in the project description (e.g., total number of schools, enrollment trends, district-wise analysis).

Visualize key findings using charts, graphs, and tables for a better understanding.

1. **Step 4: Data-Driven Insights**

Summarize the findings from your EDA. Highlight key trends and challenges.

Use your analysis to inform the allocation of the $3 billion budget to educational initiatives.

1. **Step 5: Allocation of Funds**

Based on your analysis, create a detailed budget plan that outlines how the $3 billion will be allocated to various initiatives. Consider the priorities mentioned in the project description.

1. **Step 6: Data-Driven Marketing Strategies**

Develop marketing strategies that are data-driven, considering the local context and demographics.

Ensure your strategies align with the specific initiatives you're funding.

1. **Step 7: Report Preparation**

Begin drafting the comprehensive report. The report should have the following sections:

* **Introduction**: Briefly describe the project, its objectives, and the dataset used.
* **Data Cleaning and Preprocessing**: Document the data cleaning steps.
* **Dataset Analysis**: Present your EDA findings.
* **Budget Allocation**: Explain how you propose to allocate the funds.
* **Marketing Strategies**: Describe your data-driven marketing strategies.
* **Monitoring and Evaluation**: Explain how progress and impact will be monitored.
* **Step 8: Review and Feedback**

Share the draft report with relevant stakeholders and team members for feedback.

Address any suggestions or revisions.

1. **Step 9: Finalize the Report**

Make necessary revisions based on feedback.

Ensure that the report is clear, concise, and professional in presentation.

**Data Cleaning and Preprocessing:**

* Copied the whole data into a new sheet and converted it to a table.
* Checked for duplicate
* Handled missing and blank values in **Head\_Type,School\_Type, Upgrade\_Primary\_Year, Upgrade\_Middle\_Year, Upgrade\_High\_Year, Upgrade\_High\_Sec\_Year, School\_Ownership, Construct\_Type, Bldg\_Condition, Electricity\_Source, Boundary\_Wall\_State, Care\_Giver** and replaced them with **N/A**.
* Replaced 1 as Yes and 0 as No in **Bldg\_Status, Place\_Status, Electricity, Toilets, Boundary\_Wall.**
* I've just considered **Total Area Kanal** and **Uncovered Area Kanal,** and then took their difference and converted them to Sq FT and replaced the values in the Covered Area column with these recalculated values
* Replaced blank values in **Teachers** as **No Teachers Available** and **Nonteachers** as **None**
* Replaced N/A values in **School\_Ownership** after Filtering Govt Schools from School Type to Educational Dept, as it has the majority of the values that had N/A, a total of 277, out of which 235 were govt schools, and logically, Govt Schools are owned By the Educational Dept.
* Added 2 New Columns, **Teachers Availablity and NonTeachers Availablity,** and used IF Formula to enter **Filled** if there was any teacher/non-teacher available, and filled **Vacant** if there wasn’t any teacher/non-teacher available
* There are 44 values when we filter No in Bldg\_Status, where we have some details regarding the infrastructure, but logically, if there’s no building, then how can there be other details. This is only 0.09% of the whole dataset.