

# Excel Portfolio Project



## Introduction:

The dataset provided contains school data from various locations across Pakistan. It offers valuable insights into the performance and characteristics of schools in the country. By delving into this dataset, you can better understand Pakistan's education landscape and identify improvement areas.

## Scenario:

You are a junior data analyst working with an international ed-tech company investing 3 billion dollars in Punjab, Pakistan. The company aims to leverage technology to bridge educational gaps, ensure equal opportunities for all students and improve the infrastructure and facilities. You have been given the responsibility of analysing a dataset on schools in Punjab and conducting a marketing feasibility analysis. A fund of **3 Million US dollars is allocated to the initiatives** that will contribute to improving education. The company expects you to present a comprehensive report that justifies the allocation of funds and proposes data-driven marketing strategies. The report should outline specific initiatives and projects that will **maximize the impact of the fund**, ensuring that no **money is left unattended** and all **resources are utilized** optimally.

## Deliverables:

The report you produce must include the following deliverables:

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

## Dataset Overview:

Some of the columns included in the dataset are listed below:

**School ID:** Unique identification number for each school.

**School Name:** The name of the school.

**Location:** The information on district, street, and union councils

**Enrollment:** The total number of students enrolled in the school.

**Teachers:** The number of teachers in the school

**School Upgradation Years:** The years in which the school levels were upgraded (in datasets cols are upgrade\_primary\_year, upgrade\_middle\_year, upgrade\_high\_year, upgrade\_higher\_sec\_year)

**School Gender:** The gender for which the school was initially built.

**Gender Studying:** The genders currently studying in the school

**School Head Information, School status, School medium, Year of establishment, Genders studying, Infrastructure.**

### **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

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

- Determine the total number of schools present in the dataset.
- Identify the lowest students of school\_Gender(male & female) by school level of schools established in the 2000s
- Determine the top 5 districts with the highest number of teachers and non-teachers in schools.
- Explore the school level and medium that have the lowest number of functional classrooms.
- Create a bar chart showcasing the distribution of teachers across different schools.
- Calculate the number of male and female students (School\_Gender)
- 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.
- Identify the district with the highest enrollment in primary schools since their establishment.
- Determine the district with the highest enrollment in secondary schools since their establishment.
- Find the district with the highest enrollment in higher secondary schools since their establishment.
- Create a bar chart displaying the distribution of schools according to school\_level
- Generate a bar chart representing the number of students categorized by gender (School\_Gender).
- Create a table showcasing the school\_ownership with respect to schools.
- Visualize the distribution of teachers based on their respective job posts using a bar chart.
- Create a pie chart to show the percentage of vacant and filled teaching and non-teaching posts.
- Determine the percentage of schools with satisfactory building conditions

(bldg\_condition).

- Calculate the number of schools with satisfactory security measures.
- List the names of the 5 districts with the lowest availability of drinking water facilities in schools.
- Identify the names of the 5 districts with the lowest availability of electricity in schools.
- Determine the names of the/ 5 districts with the lowest presence of boundary walls in schools.
- List the names of the 5 districts with the lowest availability of toilets in schools.

These objectives will guide your analysis and provide insights for your report. You are expected to export tables from Excel for a more comprehensive report. Additionally, feel free to explore the dataset further and present any other findings that you deem relevant for educational policymaking.

We encourage you to do extra research while constructing this report and even use other datasets to back your research further. Make sure you include your sources clearly if you do so however, and explain what other data you have used.

### **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.

Good luck, and we look forward to your findings and the report!