|  |  |
| --- | --- |
| **NAME** | Sarah Qasim |
| **MODULE NAME** | Excel |
| **DATE** | 25-Oct-2024 |

**ATOM CAMP**

**AI AND DS BOOTCAMP**

EXCEL PORTFOLIO PROJECT REPORT

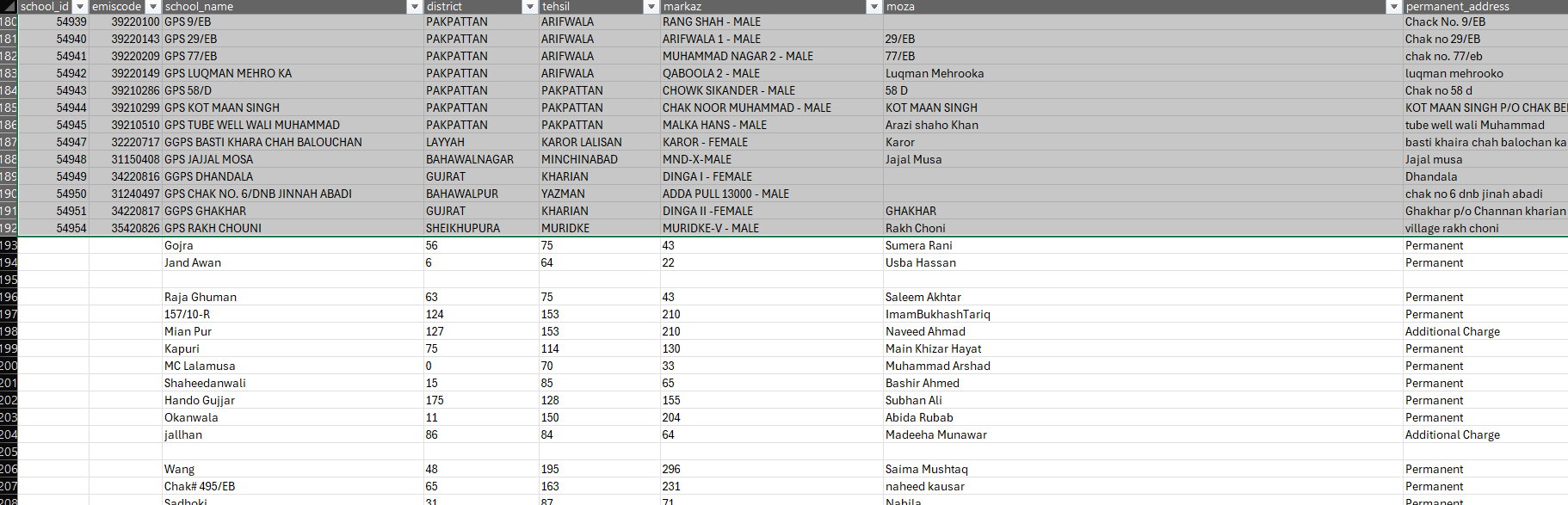
**DATA SOURCES**

1. Dataset of over 40,000 schools in Punjab which includes locations, medium of education, school levels, year of upgradation at each level, count of teachers, non-teachers and caregivers, availability of infrastructure and facilities.
2. https://stateofchildren.com/children-dataset/ : Table of “Literacy rate (10 and above)-Punjab”
3. <https://www.unicef.org/pakistan/media/3121/file/Multiple%20Indicator%20Cluster%20Survey%202017-18%20-%20Punjab.pdf>: page 213: “Primary school attendance and out of school children”
4. <https://www.worldometers.info/world-population/pakistan-population/> : Graph of “Pakistan Population (1950-2024)”

**DATA CLEANING AND MANIPULATION**

1. **Removing duplicates in the unique identifier column: school\_id**

* No duplicates were found after using conditional formatting to find duplicates
* However, there were excess values in other columns after rows finish in school\_id, removing them as the first column is our unique identifier.

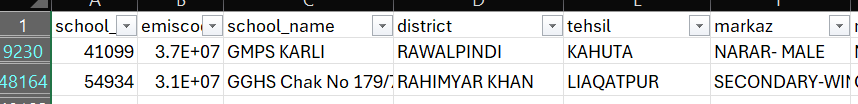


1. **Missing values**

Several columns have “blanks”, “nill”, “nil” and so forth including street name, head\_type, head\_grade, school\_ownership, construct\_type, bldg.\_condition, caregiver, teachers and non-teachers.

All of these are replaced with the phrase “no information” to maintain data consistency so when analysing the data later, it becomes easier to filter and analyse. Deleting such rows with some blanks would leave to losing valuable information in other columns of the row leading to inaccurate data analysis.

One exception was in the column “enrolment” which is numerical and only had 2 blanks cells. Imputing technique was used here instead of adding a new data type with the string “no information” for only 2 blanks.



The blanks were for districts Rawalpindi and Rahimyar Khan, so mean value of enrolment of all other schools in these districts was calculated and substituted respectively.

Rawalpindi : 211

Rahim yar khan: 232

1. **Spelling error in major category**

The category “sMosque” is replaced with “Mosque” in the column “school\_level”.

1. **Upgradation columns**

Upgradation columns for all levels (primary, middle, secondary, high\_sec) had “0”,”blanks” and “nulls” in multiple cells.

“0” and “null” were replaced with “no upgrade” as these values are answers that there has never been an upgrade while “Blanks” were replaced with “no information”.

This ensures there is uniformity and better readability of data, indicating there was no upgradation instead of arbitrary cells like “0” and “nulls”.

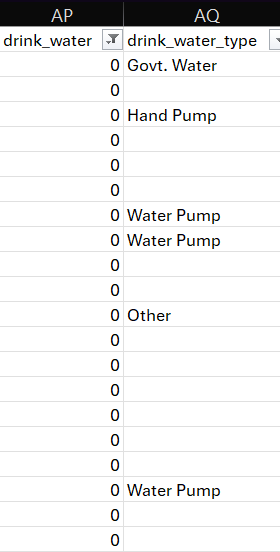
1. **Connected columns and manipulating information**

The uniform approach used in the categories (water, electricity, toilets and boundary walls) is as follows to ensure validity of collected data.

All these columns have binary answers of “0” meaning “no” and “1” meaning “yes” along with a corresponding column that is like a “if yes, then what source” column that needs to be filled if you answer “1” in the first column.

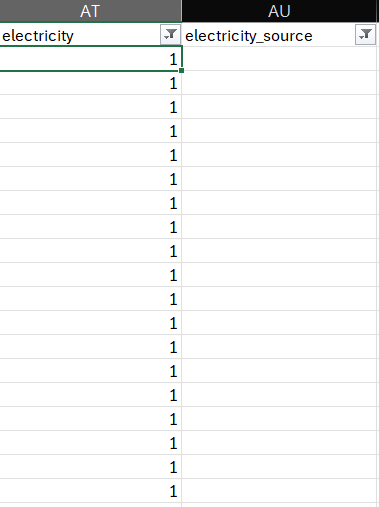
So, if they have given a source in the corresponding column, we assume there is information, so it is “1” in the first column.

Some categories give a source but say “0” in the first column, changing to “1” for all cells that say “0” and give a source name:



For “0”, and blank, replacing blanks with “no source”(water, electricity),”0” (toilet), “no boundary wall”(wall), not “no information” as there is information given in the previous column.

Some blanks correspond to “1” meaning yes there is electricity from the previous column, replace 1 with 0 as we are assuming we need a source mentioned to answer 1.



1. **Data type**

For all numerical columns, where we added “no information” or other strings, data type is changed to numeric afterwards.

**KEY TRENDS AND INSIGHTS**

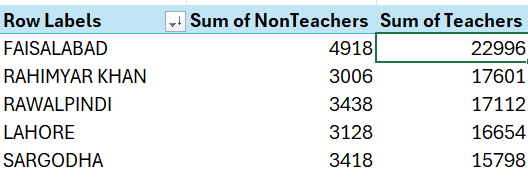
1. **A column graph to help see trends and correlations between enrolment numbers from the 2000s (recent) for both genders with different school levels.**

* We see a general downward trend in enrolment in schools made for both genders as the school level goes up but the drop in females in much higher from the primary level to higher levels.
* When comparing the two genders, the enrolment for female based schools is consistently higher at each level indicating a higher enrolment as compared to males.
* There is an outlier in the school level “middle” which ideally according to the identified trend have been higher than secondary.

I believe a better analysis, however, would occur when using the column “gender\_studying” instead as it indicates the current gender situation not the one at establishment considering we want latest stats which is why we are also using schools established in the 2000s. I will repeat this for that column to see if there is a change.

* The downward trend is much neater in the “both” column indicating a clear decrease in enrolment numbers in schools catering to both genders as the school level increases.
* Another change, in schools catering to only one gender, the secondary school level is a clear outlier in this case being higher than even the primary column. This makes the drop from secondary to higher secondary more pronounced.
* Overall, the enrolment numbers in the female schools are still higher than the male schools

1. **Top 5 districts with the highest number of teachers and non-teachers in schools.**



Now i think this data does not give an accurate picture of which districts are doing the best because some districts are larger and thus have more schools and thus more teachers and non-teachers. For e.g. Faisalabad, Lahore, Rawalpindi and Rahimyar khan are just the largest districts in the province.

I am changing this to get two more results that I feel will paint a better overall picture:

* **Ratio of sum of teachers in the district to number of schools in that district giving us number of teachers per school in each district.**

**Formula:**

*=GETPIVOTDATA("Sum of ",$B$21,"district",B22)/GETPIVOTDATA("Count of school\_id",$B$21,"district",B22)*

Using conditional formatting to highlight the “top 5” values.

**A screenshot of a table

Description automatically generated**

**Highest: Lahore, Faisalabad, Rawalpindi, Khanewal, Sahiwal**

Lahore moves up to the first position by a large margin from the others.

While smaller cities like Khanewal and Sahiwal despite having lower overall values of teachers perform way better now that number of schools are considered!

A screenshot of a table

Description automatically generated

**Lowest: Bahawalnagar, D.G. Khan, Rahim yar khan, Layyah, Bhakkar**

Now interestingly, Rahimyar Khan is in the bottom 3 in this perspective from top 5 in the previous analysis of the same question!

Thus, this shows a bigger and more analytic picture that due to the size of Rahimyar khan the sum of teachers was very high however, the number of teachers per schools performs poorly indicating a serious lack of staff resources.

* **Ratio of sum of non-teachers in the district to number of schools in that district giving us number of non-teachers per school in each district.**

Conditional formatting:

**Green : top 5**

**Red: bottom 5**

A screenshot of a data

Description automatically generated

**Top 5: Lahore, Faisalabad, Attock, Khanewal, Sahiwal**

**Bottom 5 : Rajanpur, Pakpattan, D.G Khan, Mianwali, Bhakkar**

Most districts stay around 1 meaning 1 non-teacher per school thus most districts in Punjab are heavily understaffed in terms of non-teachers, exposing a large gap in learning facilities available in the province.

The ratio of non-teachers is much lower than that of teachers, with a range 0.9-2.8 compared to 6-14 for teachers per school.

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

A screenshot of a computer

Description automatically generated

**Lowest: Mosque, English**

The value of functional classes in mosques in Urdu is more than 13 times this value indicating the general idea that most mosque education caters to students who only speak Urdu from a specific subset of society.

**Highest: Primary, Urdu**

Reflects further upon our previous result of the highest enrolment being in primary level schools which leads to the most functional classes as a number not a ratio to be dominated by primary education.

Urdu is the most dominant medium in Punjab schools and thus automatically leads the number of functional classrooms as well.

A graph of different colored bars

Description automatically generated

**Overall:**

* There is a downward trend of functional classes as the school level increases across all mediums, this consolidates our earlier trend analysis of enrolled children at each level.
* Urdu medium schools dominate at all levels (except an outlier in higher secondary level but by a small margin) while English medium schools are the lowest.

1. **The number of male and female students (School\_Gender)**

I have added pie charts to compare gender division when school was established (school\_gender) and gender division now (gender\_studying) to notice the shift over time.

Significant decrease in both female and male from established is converted to an increase in the number of schools that cater to both genders instead.

***Further analysis:***

While this may look like a positive thing as a shift in culture to cater to gender mixing, it could potentially lead to parents, especially of girls, not wanted to send their children to schools that cater to both genders leading to a decrease in female enrolment.

This could also be due to a lack of resources and funding due to which schools have been combined for both genders instead of building more schools for to cater to a single gender.

To investigate these potential reasons, I have looked at number of students enrolled per school divided by gender currently studying there:

A close-up of a number

Description automatically generated

As suspected, despite 63% of schools catering to both genders, the number of students per school is less than half of the number of students per school in single gender schools.

This confirms my concern that increasing the number of co-education schools has unfortunately led to people being reluctant to send their children (especially girls in Pakistani culture) to these schools.

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.**

Adding a pie chart to display the two categories better visually:

* **Division according to medium**
* Urdu medium in rural overshadows all other quantities
* A downward trend in both rural and urban with Urdu being the highest and English being the lowest.
* Trend stays consistent throughout in both locations, however the difference between Urdu and English is much smaller in urban cities as compared to rural where Urdu dominates with a staggering 72% of the total rural schools.

A graph with numbers and text

Description automatically generated

**JUSTIFICATION OF ALLOCATION OF FUNDS**

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

1. **School level upgradation**

**Identifying problem:**

* ***Number of schools updated each year by school level in the 2000s***

* A clear downward trend in number of schools updated per year at each level (except secondary) with all levels except secondary reaching their lowest ever recorded number of updated schools in the latest year.
* Most overall upgrades occurred in middle schools despite being much smaller in total number of schools than primary indicating a higher ratio of updated schools.
* An outlier is an upward spike in the years 2011-2012 for all levels except primary.
* ***Do districts with more updates have better facilities?***
* A clear correlation is seen between the schools that were recently updated and the schools with available drinking water and electricity.
* Only one district “Chiniot” had some schools which were updated but did not have drinking water or electricity.
* ***Do districts with more updates have better infrastructure?***
* A similar positive correlation can be seen here between schools that were recently updated and schools with optimal infrastructure (bldg\_status and boundary wall).
* Only two districts “Chiniot” and “Bahawalnagar” had schools which were updated but did not improve infrastructure.

**Justifying fund allocation:**

The most recent year reaching the lowest level of upgradation in most school levels indicates an urgent need for upgradation at all levels. The spike in 2011-2012 shows how a targeted effort like this can boost the school upgradation levels.

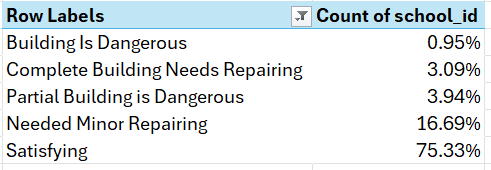
Especially targeted towards primary level as it has the highest percentage of enrolment across levels but has the lowest number of schools updated in the recent number of years meaning so a large number of students across Punjab are studying in schools that have never been upgraded.

The positive correlation between improved infrastructure and facilities and school upgradation shows how this is vital step to directly impact the quality of education students receive.

1. **Infrastructure**

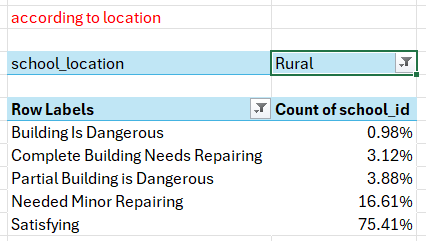
**Identifying problem:**

* ***Schools and building conditions***



Overall, building condition is “satisfying” in around 75% of schools surveyed showing a positive trend.

* ***Dividing by location: rural and urban***

**A screenshot of a computer

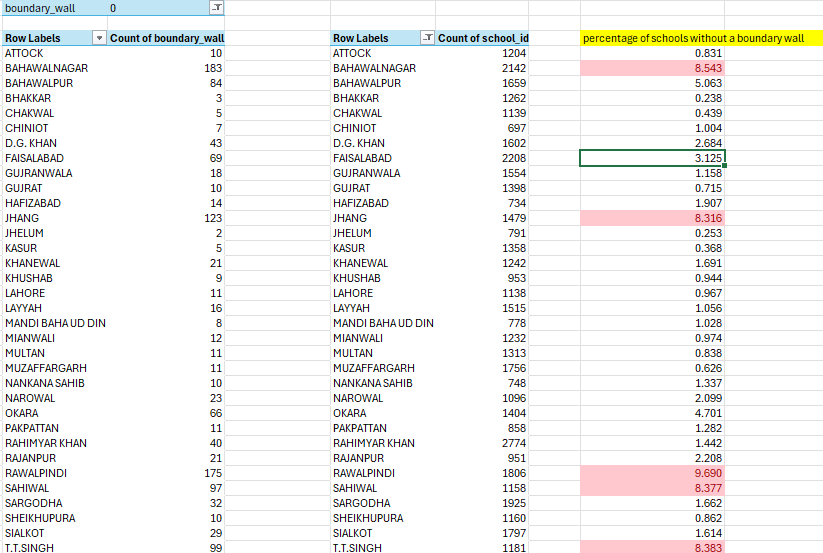
Description automatically generated**

Trend stays the same when divided in rural and urban areas indicating that the building conditions are generally satisfying in both less developed and more developed regions of the province which means we can concentrate more funds on other infrastructure areas.

* ***Percentage of schools without a boundary wall in each district***

Using the same method as before and dividing count of schools without a boundary wall (filter = “0”) by number of schools in the district and multiply by 100.

Using Conditional formatting to highlight the “top 5” districts with highest percentage.



**Rawalpindi, Bahawalnagar, T.T Singh, Sahiwal, Jhang**

* ***Percentage of schools without toilets in each district.***

Using the same method as before and dividing count of schools without toilets (filter = “0”) by number of schools in the district and multiply by 100.

Using Conditional formatting to highlight the “top 5” districts with highest percentage.

A screenshot of a computer

Description automatically generated

**Bahawalnagar, Rajanpur, D.G. Khan, Rahimyar Khan, Hafizabad**

**Justifying fund allocation:**

The clear lack of basic infrastructure like toilets and boundary walls in districts with a large number of schools like Bahawalnagar, Rawalpindi and Rahimyar Khan indicates the dire need to use the allocated 5M rupees for improving the situation.

The funds can be concentrated on these areas to improve student and teacher hygiene and create a safer environment for students especially younger children at primary level.

This will in turn also increase the number of parents willing to send their children to these schools improving enrolment numbers.

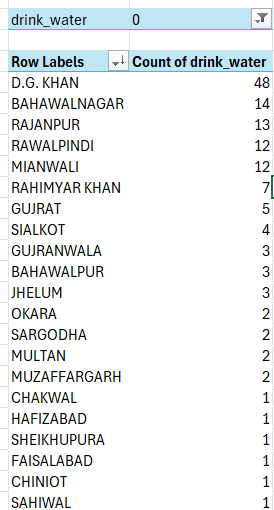
1. **Facilities**

**Identifying problem:**

* ***Districts with highest non-availability of drinking water***

**Unavailable:** count of 0s

**Highest:** sort in descending order

******

* Only 5 out of the 36 districts have more than 10 schools without drinking water.
* Most districts only have 1-3 reported school(s) without drinking water.
* ***Districts with lowest availability of electricity***

Using the same method as before and dividing count of schools without electricity (filter = “0”) by number of schools in the district and multiply by 100.

Using Conditional formatting to highlight the “top 5” districts with highest percentage.

A screenshot of a computer

Description automatically generated

**Rajanpur, Bahawalpur, Muzaffargarh, Rahimyar Khan, Bahawalnagar**

**Justifying fund allocation:**

Allocating the smallest portion of total funds for improving facilities is a good strategy as only a limited number of schools lack drinking water and electricity across the dataset and thus the majority of the funds can be dedicated to more pressing and widespread issues.

1. **More middle and high schools in rural areas with separate schools for both genders**

**Justifying fund allocation:**

* ***Rural distribution***

Most schools are in the rural parts of the province

* Urdu medium in rural overshadows all other quantities
* Trend stays consistent throughout in both locations, however the difference between Urdu and English is much smaller in urban cities as compared to rural where Urdu dominates with a staggering 72% of the total rural schools.

A graph with numbers and text

Description automatically generated

* ***Shift to mixed gender schools and how that effects female enrolment***

While this may look like a positive thing as a shift in culture to cater to gender mixing, it could potentially lead to parents, especially of girls, not wanted to send their children to schools that cater to both genders leading to a decrease in female enrolment.

This could also be due to a lack of resources and funding due to which schools have been combined for both genders instead of building more schools for to cater to a single gender.

To investigate these potential reasons, we now look at the number of students enrolled per school divided by gender currently studying there:

A close-up of a number

Description automatically generated

As suspected, despite 63% of schools catering to both genders, the number of students per school is less than half of the number of students per school in single gender schools.

This confirms the concern that increasing the number of co-education schools has unfortunately led to people being reluctant to send their children (especially girls in Pakistani culture) to these schools thus separate gender schools are needed to increase female enrolment.

* ***Why focus on medium and secondary level schools in rural areas:***
* Downward trend from primary to higher secondary.
* A huge drop in percentage of schools working at primary level to those working at secondary level.
* Development of schools in medium and secondary level is the first step to encourage people to send children to school even after primary school. The next step could then focus on higher secondary level schools, but the urgent need is to localize funds effectively on this first step.

**External data:** https://stateofchildren.com/children-dataset/

A screenshot of a computer

Description automatically generated

To emphasize why the largest portion of the budget is allocated in this section, we have included the literacy rate of children in Punjab from an external dataset, which highlights the blaringly low number of less than 50% literacy in females in rural areas.

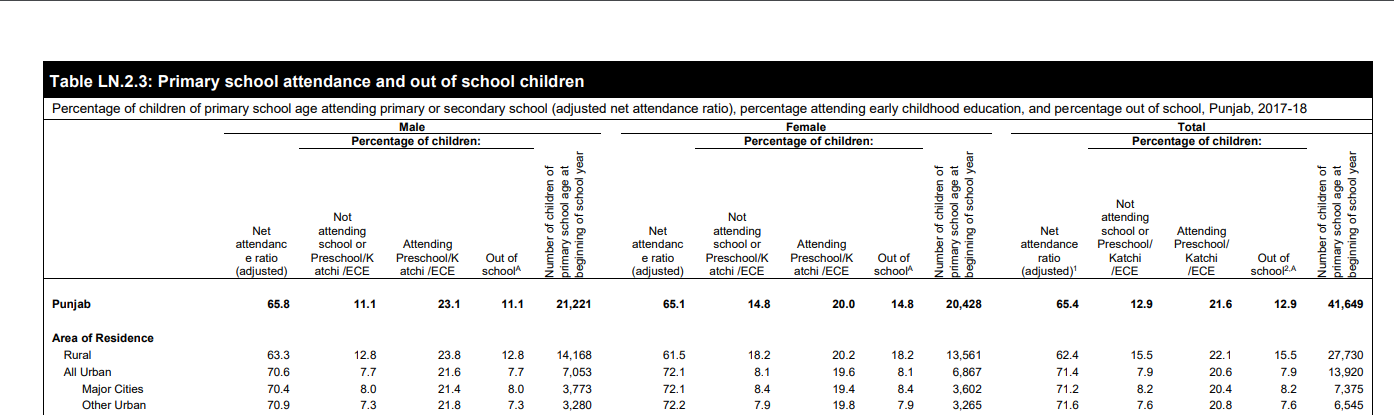
* **Is work for this initiative being done already?**

Despite the already stated need to focus on developing this area, the dataset shows a sharp decline in the last 20 years with only a few secondary level schools were established recently while no new middle level schools have been established since 2014.

1. **New primary level schools**

**Justifying fund allocation:**

* ***Primary school attendance and out of school children at primary school age***



**External source:** [**https://www.unicef.org/pakistan/media/3121/file/Multiple%20Indicator%20Cluster%20Survey%202017-18%20-%20Punjab.pdf**](https://www.unicef.org/pakistan/media/3121/file/Multiple%20Indicator%20Cluster%20Survey%202017-18%20-%20Punjab.pdf) **(**page 213) of the pdf

Zooming in to the Punjab portion:

**A screenshot of a computer

Description automatically generated**

12.9% of children of primary school age do not attend school, meaning they are not even receiving the most basic education needed for basic literacy.

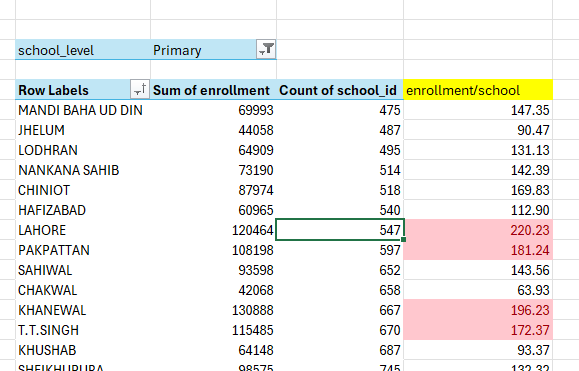
* ***Enrolment per number of schools in each district for primary level schools***

A screenshot of a data

Description automatically generated

**Lahore, Khanewal, Faisalabad, Pak Pattan, T.T Singh**

Sorting “count of school id” from smallest to largest:



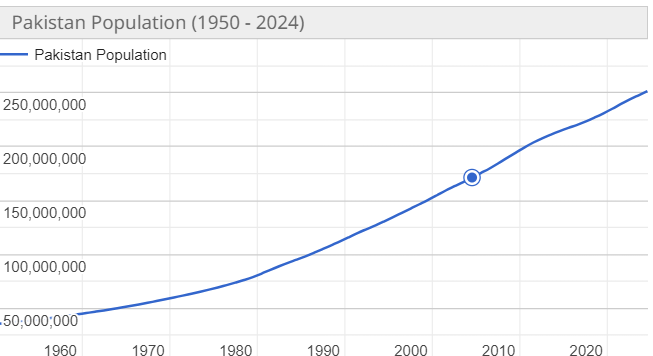
A general trend: 4/5 districts with the highest enrolment numbers/school are in the top 3rd of districts with lowest number of schools indicating there is a clear trend of these schools being overcrowded.

* ***Find teacher: student ratio in each district***

High student-to-teacher ratio at the primary school level in most districts.

This is concerning as young children especially need more one on one attention, care and engagement from their teachers. Constructing more primary schools especially in rural areas is vital to thus spread the large number of students enrolled and reduce over crowding.

* **Is work for this initiative being done already?**

****

**External Source:** [**https://www.worldometers.info/world-population/pakistan-population/**](https://www.worldometers.info/world-population/pakistan-population/)

Despite the exponential increase in population (and consequently children under 5), the number of primary schools established for these children has continued to decrease since 2013 thus justifying the allocation of a large portion of funds for this initiative.

**MARKETING STRATEGIES**

***Promoting female enrolment***

Promote safety of young girls in well-built, separate gender schools with water and electricity facilities to help ease traditional parents into allowing their daughters to continue school after primary level into middle and secondary level.

***Promoting a better future***

Provide real-life examples of individuals who went from under-privileged backgrounds (similar to the audience) to becoming successful people though the power of education.

This material incentive is likely to encourage people to promote higher level of education in their own children.

***Increased employment opportunity***

Building new schools at all levels would lead to hiring of teachers and non-teachers from the local area thus promoting employment in skilled, educated labour.

Improving on infrastructure like consolidating walls or installing water and electricity sources would also lead to employment of skilled, manual labour as well.

This would help promote these incentives as not just a positive impact for children but for all people in the area.

***Get on board with local leaders***

In rural areas, people are closely knit and are most likely to listen to local people they know and trust especially local leaders.

Encouragement and support from these leaders would increase the chance of a positive response from the general people.

Get local leaders on board by providing on ground data which shows the correlation between increased number of schools and development in an area.

**BUDGET PLAN**

|  |  |  |  |
| --- | --- | --- | --- |
| **Initiative** | **Budget** | **General**  **Breakdown** | **Specific Areas** |
| **School level upgradation** | 10 M | * Primary: **3M** * Middle: **3M** * Secondary: **3M** * Higher: **1M**   (least number of schools at this level) | * Chiniot * Rahimyar Khan * Bahawalnagar |
| **Infrastructure** | 5M | * Boundary Wall: **3.5M** * Toilets: **1.5M**   (lower ratio of districts without toilets) | * Bahawalnagar * Rawalpindi * Rajanpur |
| **Facilities** | 2M | * Drinking Water: **1M** * Electricity: **1M** | * D.G. Khan * Rajanpur * Bahawalpur |
| **More middle and high schools in rural areas with separate schools for both genders** | 20M | * Building Middle schools (f and m): **10M** * Building High schools (f and m): **10M** | * Rural Regions |
| **New primary level school** | 10M | * Building primary level schools: **10M** | * Lahore * Khanewal * Pakpattan |

**SUMMARY**

The analysis of schools in Punjab reveals that most schools are located in rural areas. Enrolment declines from primary to middle and high school in both urban and rural areas for both genders while enrolment of females continues to be greater than male enrolment. A concerning result is of low teacher-student ratio and overcrowding at primary level in some of the biggest districts in the province. Despite these challenges, most schools in both rural and urban areas report satisfactory building conditions, and a low percentage of non-availability of electricity and drinking water, indicating existence presence of basic infrastructure.

**EXCEL DASHBOARD**

**(UPLOADED AS WELL)**

**A close-up of a graph

Description automatically generated**