ASSIGNMENT # 2- GAME DATASET REPORT

Hassan Aamir

BSCS13D 453976

1. Description and Understanding of the Data

1.1 Overview of the Dataset

This dataset contains information about the medal count of nine countries and four groups over various games. It also has prestige score belonging to each game. There is also the total medals achievable per game.

Derived Analytical Variables:

These are the fields we created to assist in exploration and explanation.

1. prestige_quartile: Games divided into four bins on basis of prestige.
2. pk_share: The share of total medals per game for pakistan
3. groups_sum: Sum of all group medals per game
4. countries_sum: Sum of all country medals per game

1.2 Data Cleaning & Preparation

The raw data required a number of steps in terms of preprocessing:

1. **Group1 Cleaning:** We cleaned an inconsistent entry in Group1 having a text value.
2. **Type Conversion:** Group1 was successfully converted to numeric data type

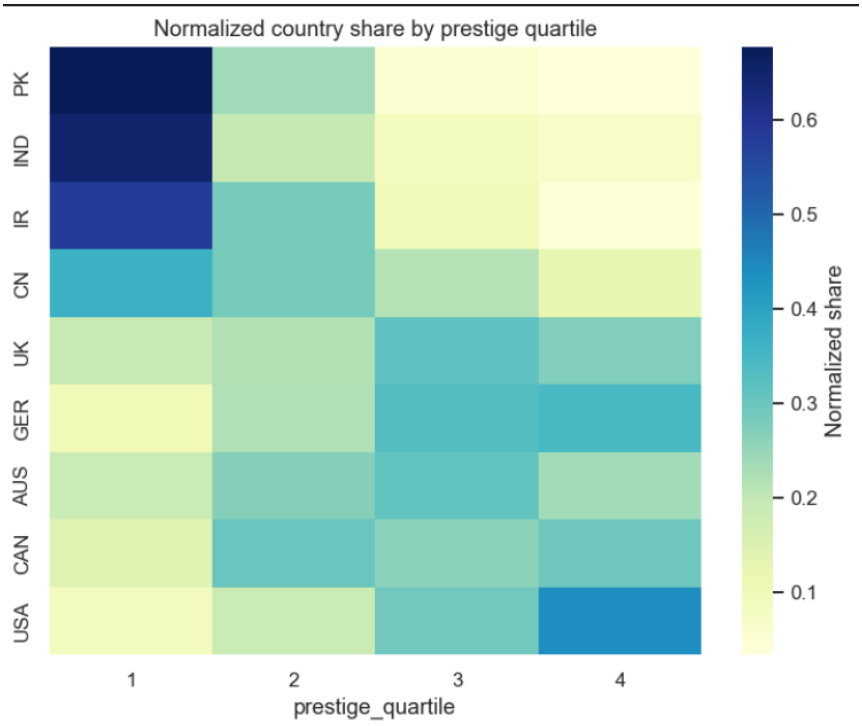
2. Exploratory Data Analysis (EDA)

The EDA phase was about the discovery of patterns, distributions, and relationships in the data without making any causal claims. This section addresses the question: "What happened?"

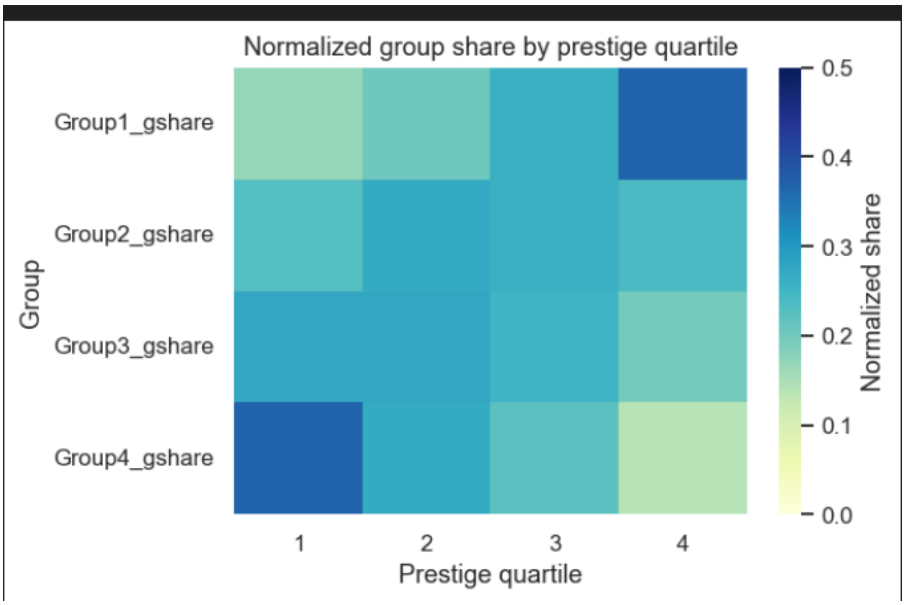
2.1 Prestige, Group Performance & Country Performance

Objective: To relate game prestige with Group Contributions (Group1-4) and to the medal count of Individual countries to find an overall trend and pattern.

2.1.1 Normalized Country/group Share Heatmap



Created a heatmap showing each country's normalized share of medals across prestige quartiles (row-normalized to sum to 1).



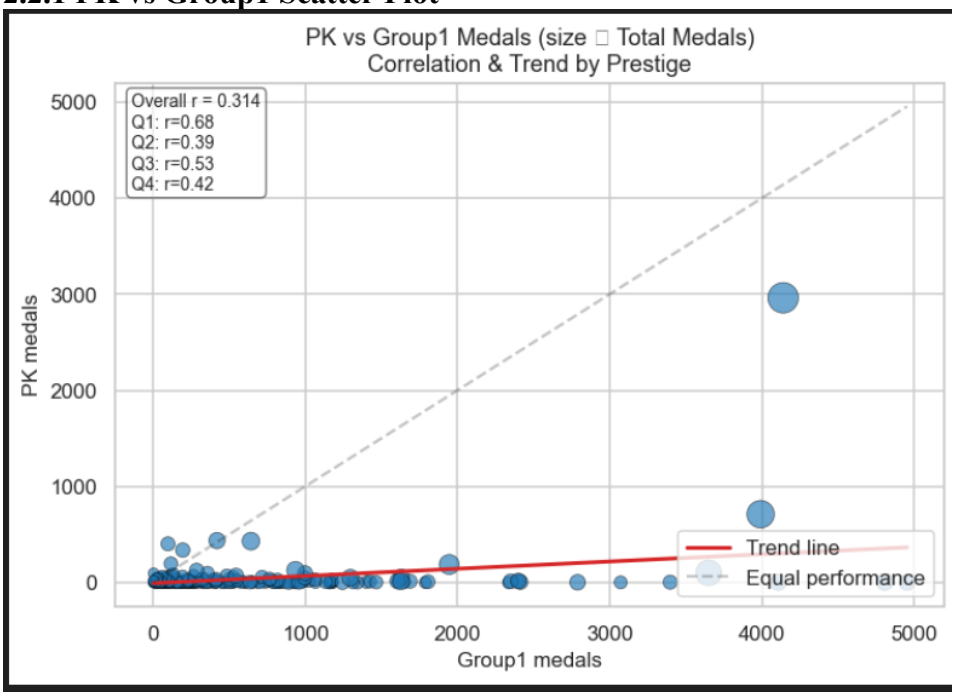
Key Findings:

1. Pakistan has most of its share in medals in Low prestige games.
2. Other countries/group concentrate on other prestige games, indicating different strategy according to the prestige of the game
3. The heatmap shows which countries/group dominate which prestige tiers

2.2 Pakistan vs Group1 Performance

Objective: To directly compare the medal performance of Pakistan against Group1 to determine if they track together or diverge under different conditions..

2.2.1 PK vs Group1 Scatter Plot



Scatter plot with marker size proportional to total medals, with regression line, including correlation metrics by prestige quartile.

Key Findings:

1. Pakistan is moderately positively correlated with Group1 medals.
2. The strength of associations varies across prestige quartiles, suggesting context-contingent relationships.
3. Points below the diagonal line represent games when Group1 outperformed PK
4. Regression line shows the general trend

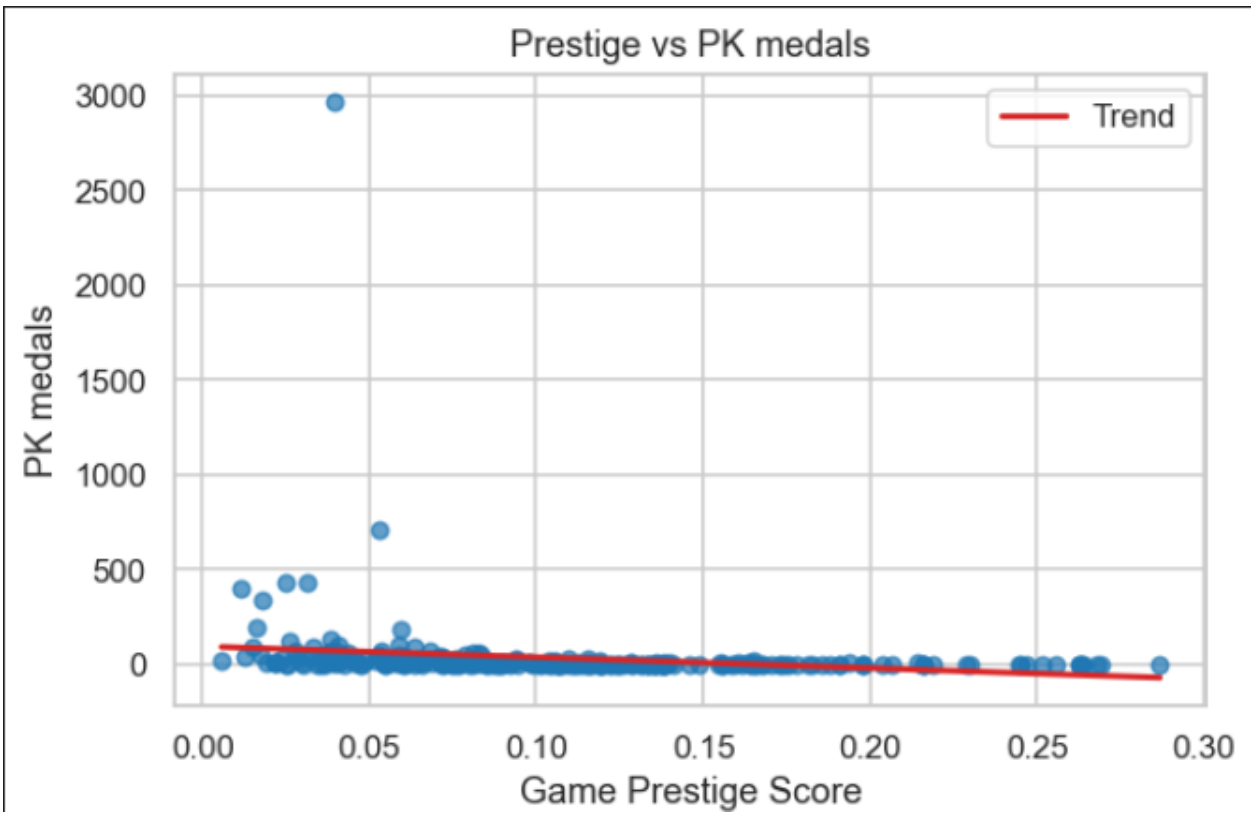
Explanation:

1. There is a moderate positive correlation between Group1 versus PK medals, which means that with an increase in the performance of Group1, the performance of Pakistan also increases on medal winning.
2. However, it is not a perfect strength association. That suggests other factors or groups influence the medal outcome of PK.

2.3 Pakistan Low-Prestige Hypothesis

Objective: Investigate the hypothesis that Pakistan does better in low-prestige games by comparing the distributions of prestige scores between PK wins and non-wins.

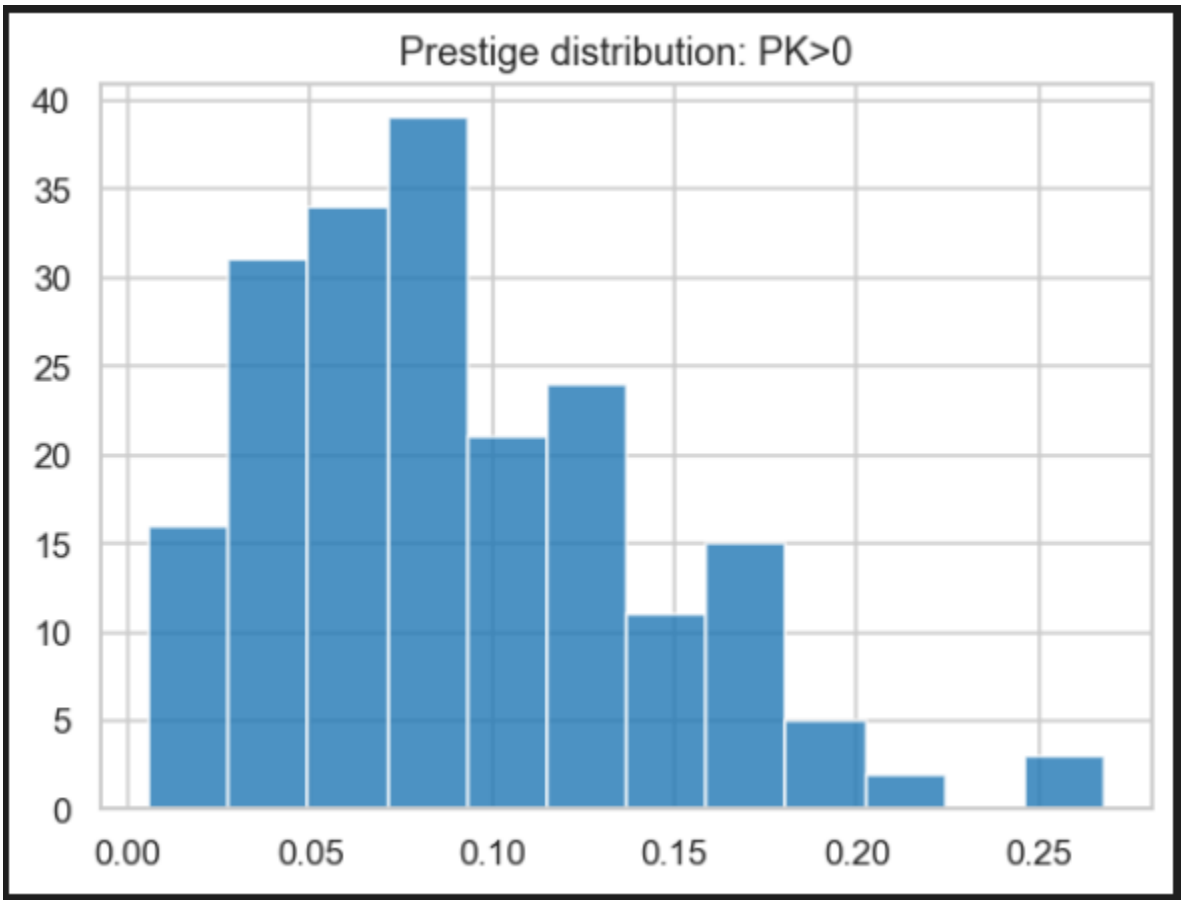
2.3.1 Prestige vs PK Medals



Key Findings:

1. As prestige rises, medal count of Pakistan declines.
2. This shows Pakistan is not interested or has disadvantage in higher prestige games.

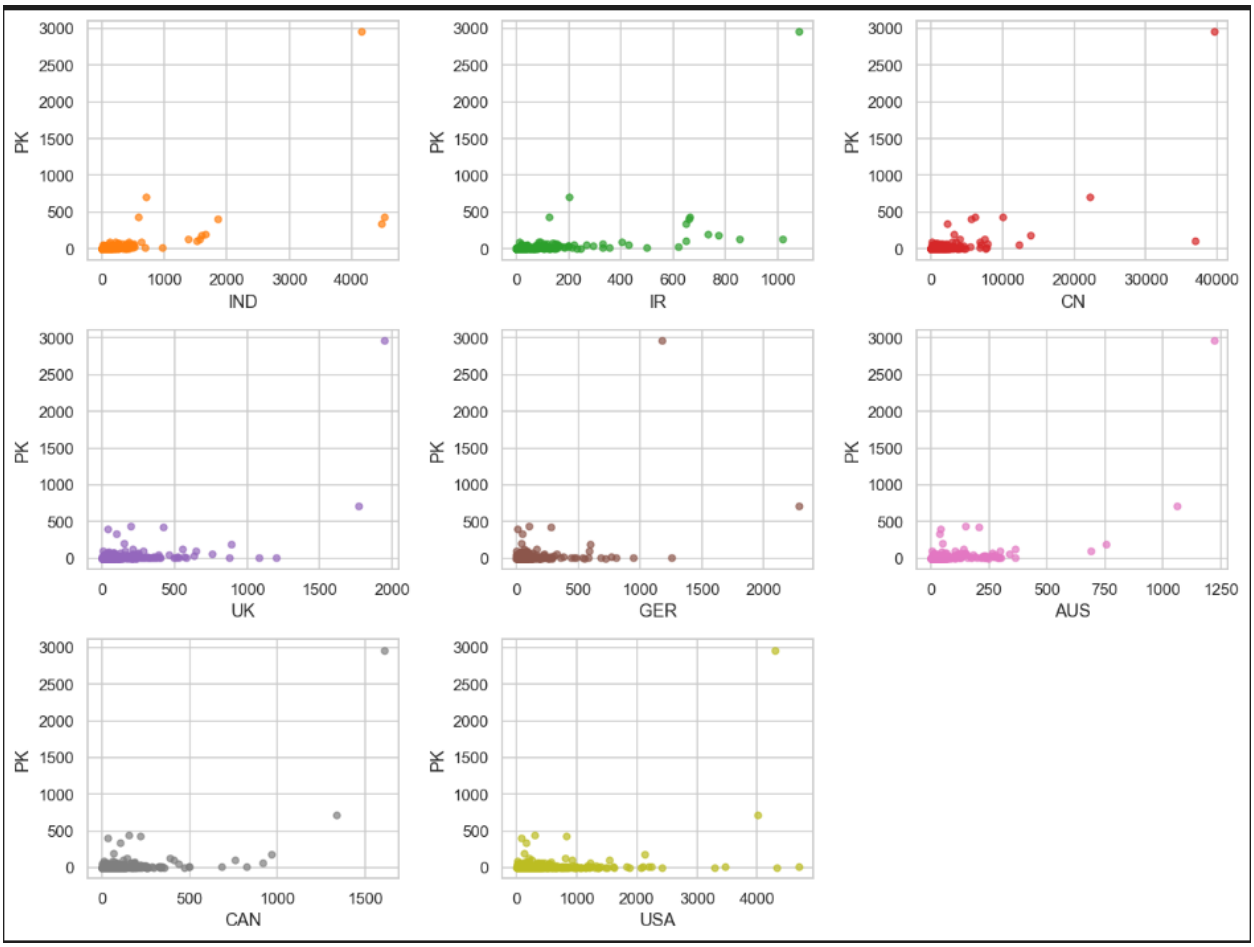
2.3.2 Prestige Distribution for when PK wins



Key Findings:

1. The wins of Pakistan is skewed towards lower prestige games.

2.3.3 Pakistan vs Each Country



Key Findings:

1. On the whole, Pakistan wins fewer medals than almost any other country.

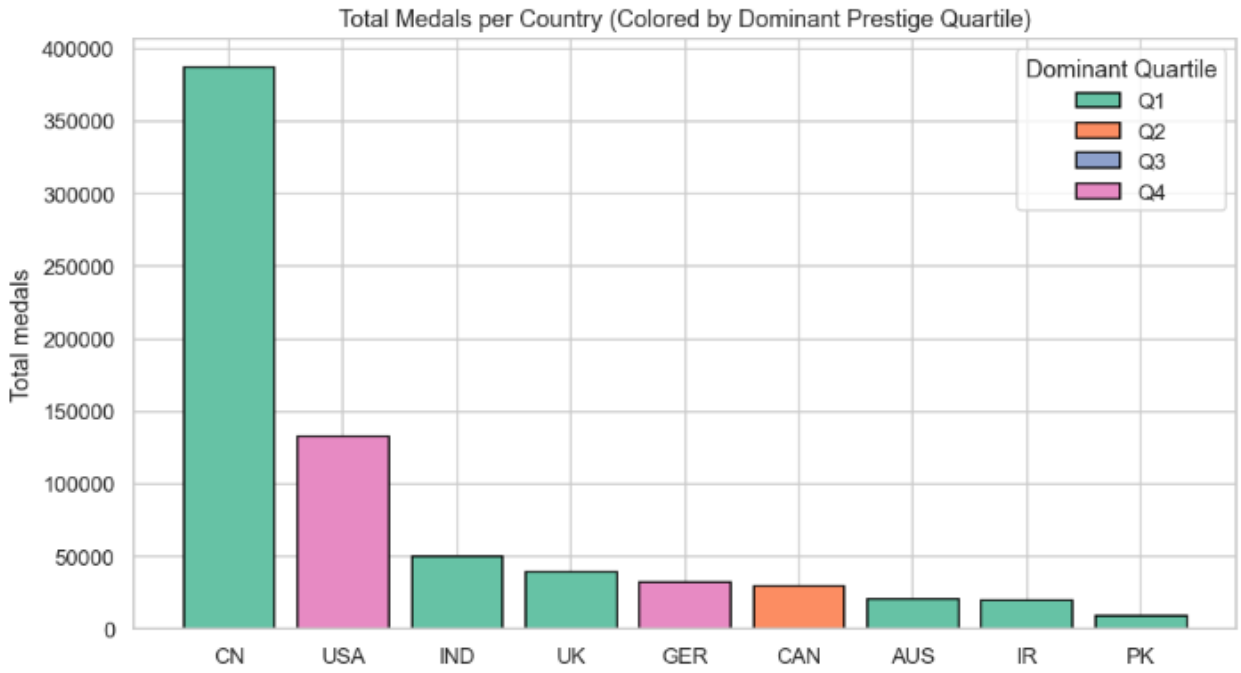
3. Explanatory Data Analysis

In this section we are answering the question: "Why did it happen and under what conditions?"

3.1 Prestige, Group Performance & Country Performance

Argument: Every country has a different prestige profile that shows strategic strengths and competitive positioning.

3.1.1 Country Dominance Report

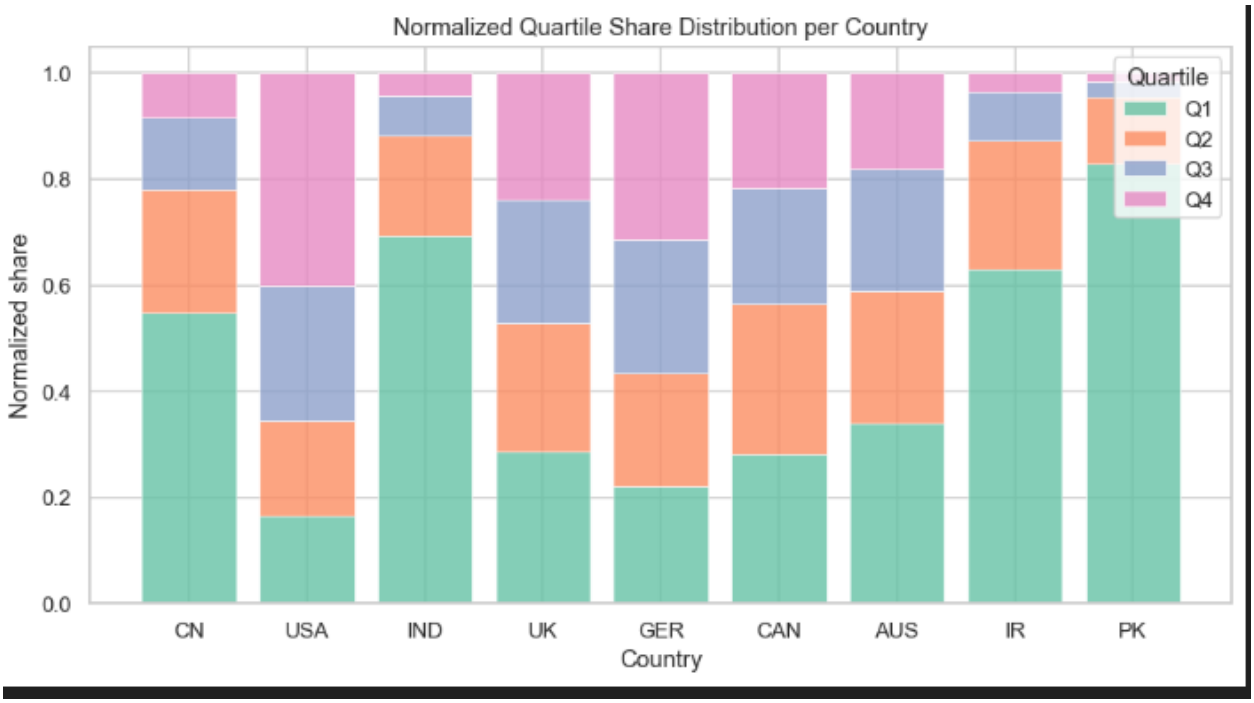


We developed an overall report that, for each country, showed the total of medals, and the share of those medals from each prestige quartile; then identified the "dominant quartile"-the prestige level at which each country wins the most medals.

Key Findings:

1. We see the medal counts of each country with the prestige quartile they dominated, higher quartile being the range of the higher prestige level.

3.1.2 Distribution of Normalized Quartile Share by Country



Stacked bar chart displaying the breakdown of each country's medals across prestige quartiles (each bar sums to 1).

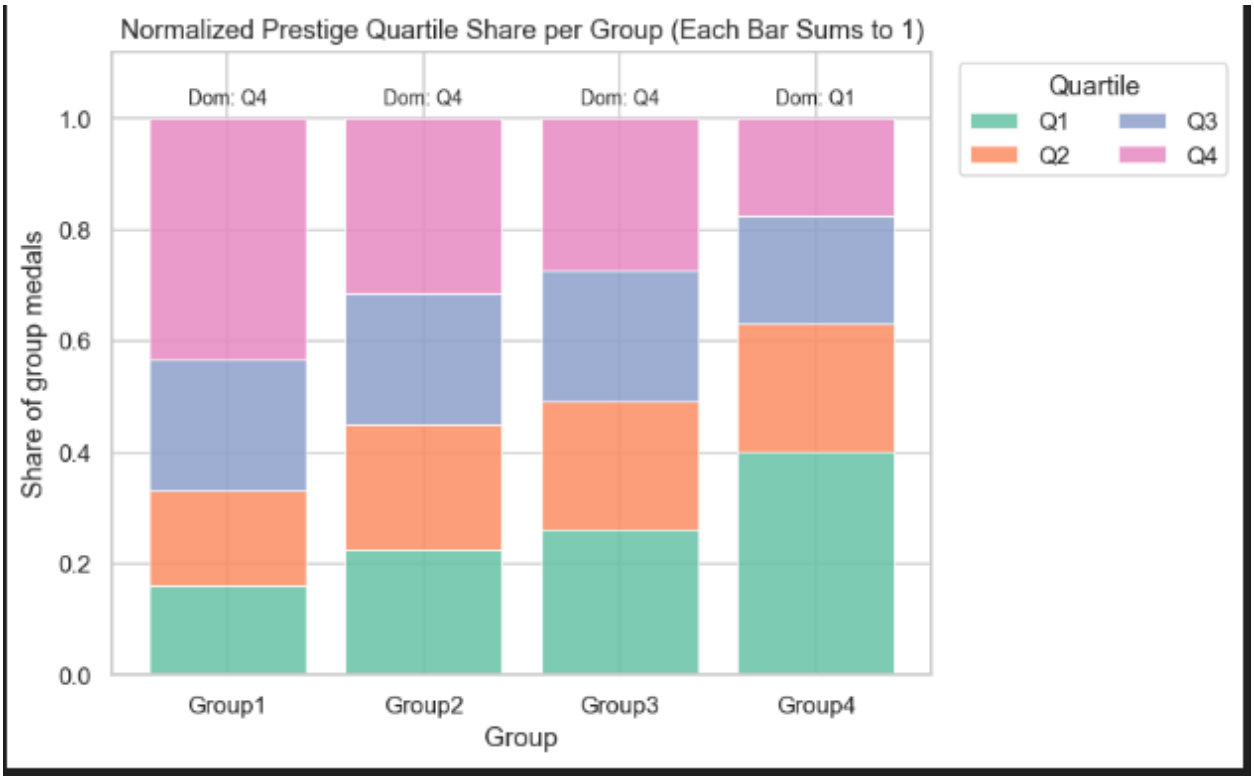
Interpretation:

1. Each segment shows the ratio of the country's medals belonging to a particular quartile
2. Tall lower segments Q1 reflect concentration in low-prestige games
3. Large dispersals indicate wider range performance of prestige.
4. USA shows strong Q3-Q4 concentration, while PK shows Q1 concentration.

Key Insight: Each country thus has a prestige sweet spot, wherein they win more medals; this being a reflection of strategic positioning, resource allocation, or competitive advantage.

3.1.3 Group Quartile Share Distribution

Gr



Similar analysis for Group1–Group4 showing the share of each group's total medals that come from each prestige quartile.

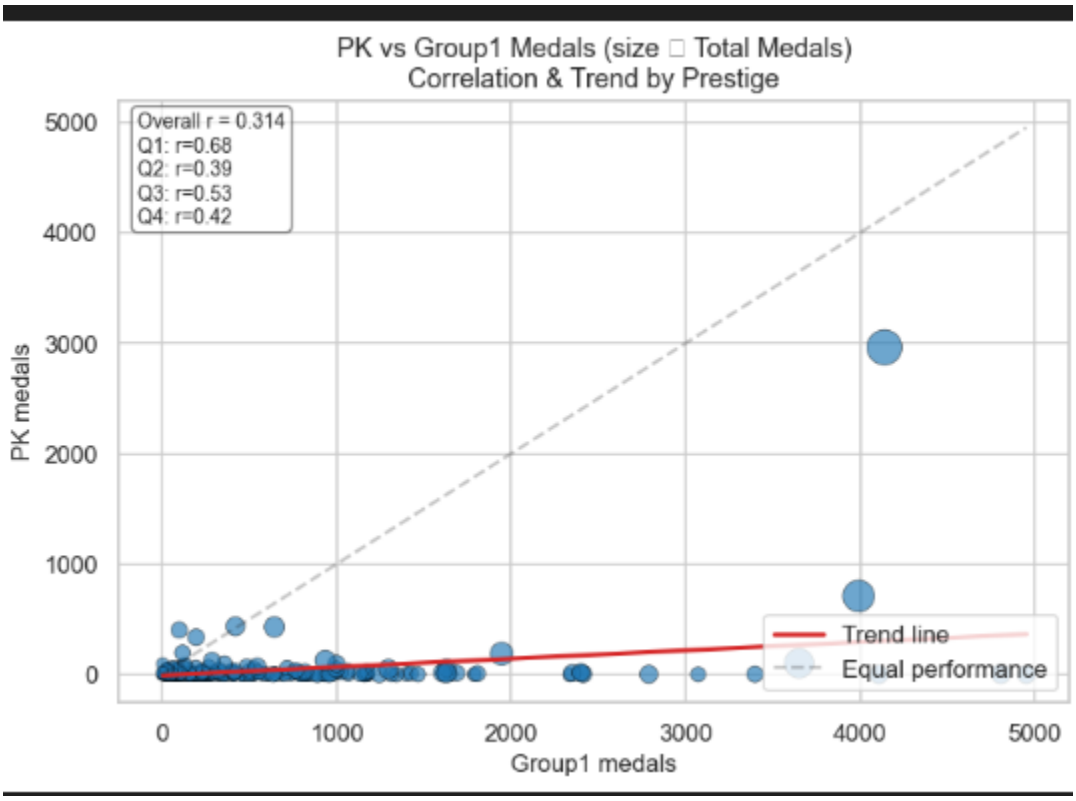
Key Findings:

1. In group 1, 2 and 3 the dominant quartile is Q4.
2. In group1, the dominant quartile is Q1
3. This suggests different strategic focuses or competitive capabilities across groups for playing in the games.

3.2 Pakistan vs Group1 Performance

Argument: Pakistan's performance, relative to Group1, varies systematically by game prestige, and understanding this ratio reveals competitive windows.

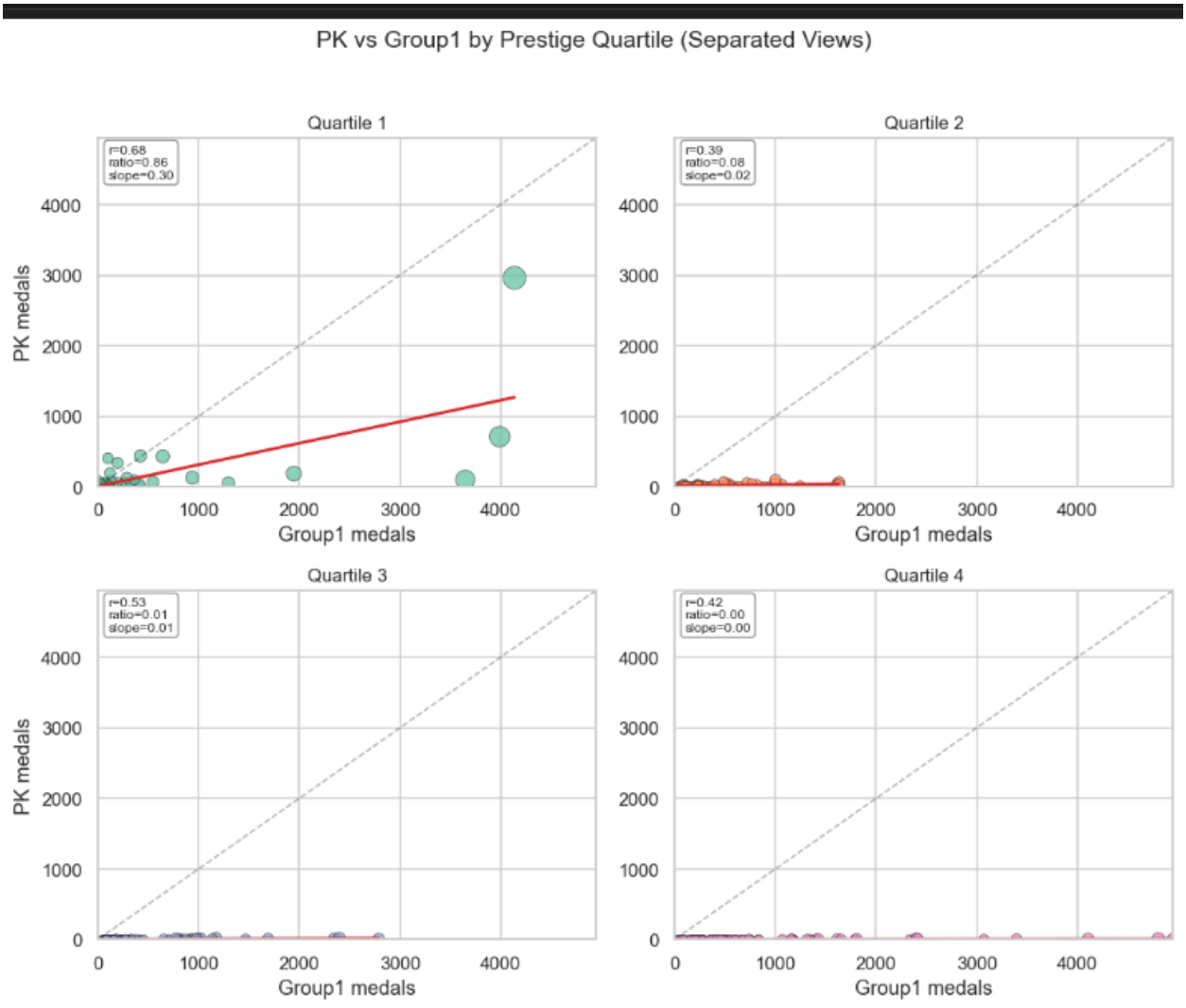
3.2.1 PK vs Group1 Scatter



Interpretation:

The Pk's medals are somewhat dependant on the medals in group1. However, large scatter about the regression line indicates that there other factors that affect PK's performance rather than group1.

3.2.2 Small-Multiples Scatter: PK vs Group1 by Prestige Quartile



Interpretation:

In Quartile 1 (Low Prestige): Higher average ratio PK/Group1 mean and stronger correlation. In Quartile 4 (High Prestige): Lower average ratio PK/Group1and weaker or negative correlation. What is more, the slope is changing across quartiles, thus reflecting prestige-dependent strength of the relationship.

Explanatory Evidence:

1. There is a moderate correlation of Group1 with PK medals across all the games
2. The relationship is not perfect in strength. This means other factors or groups are contributing to PK's medals.
3. Quartile-stratified analysis shows this association varies by prestige level; PK is possibly more competitive in the lower-prestige contexts.

4. Narrative And Story

4.1 Prestige among different Countries

Not all countries compete equally across all prestige levels. On the contrary, every country has a different prestige profile reflecting strategic strengths, resource allocation, or competitive advantages.

Supporting Evidence:

1. Prestige Quartile Analysis reveals that there is systematic variation in the distribution of medals by level of prestige.
2. The Country dominance report shows that each company has its own strategy to reach the the top. They dominate different quartiles of prestige.
3. Normalized Stacked Bars show allocation profiles, not absolute scale.

4.2 Strategic Position of Pakistan

Finding: It specializes in lower-prestige games.

Supporting Evidence:

1. The dominant quartile of Pakistan is Q1.
2. Prestige score relates negatively to PK medals.
3. Distribution histogram of prestige for PK wins is skewed to lower prestige ranges.

Interpretation:

We see that Pakistan's resource allocation may be on winnable competitions. Or this could reflect competitive constraints i.e. difficulty competing in high-prestige events.

4.3 The PK-Group1 Relationship

Finding: The performance of Pakistan relates to Group1 only moderately. The relation is also dependant on the prestige level of the games.

Supporting Evidence:

1. Overall, correlation coefficient is moderate - neither strong, nor weak.
2. Quartile-stratified analysis shows varying strengths of correlation by prestige level
3. Mean ratio PK/Group1, declines with prestige

Interpretation:

Every time Group1 does well, Pakistan is doing well, but not always. Other factors include Group2-4 and country-specific variables. There maybe other game characteristics characteristics that may affect PK performance. So, the relationship is context-dependent: stronger in low-prestige games and weaker in high-prestige games. This implies that PK's competitive advantage might be linked to the same conditions that favor Group1 in low-prestige contexts.

4.4 Group Structure and Prestige

Finding: Groups 1-3 dominate higher-prestige games; Group 4 dominates lower-prestige games.

Supporting Evidence:

1. Group quartile share analysis shows Q4 dominated by Groups 1 to 3.
2. Group4 shows the dominance of Q1 similar to the pattern seen in Pakistan.
3. Normalized stacked bars reveal clear differences in the allocation.

Interpretation:

Group structure reflects underlying competitive capabilities or strategic focuses. Moreover, PK's prestige profile and the Group4 profile are aligned. This shows that may be PK and group1 have some shared limitations.