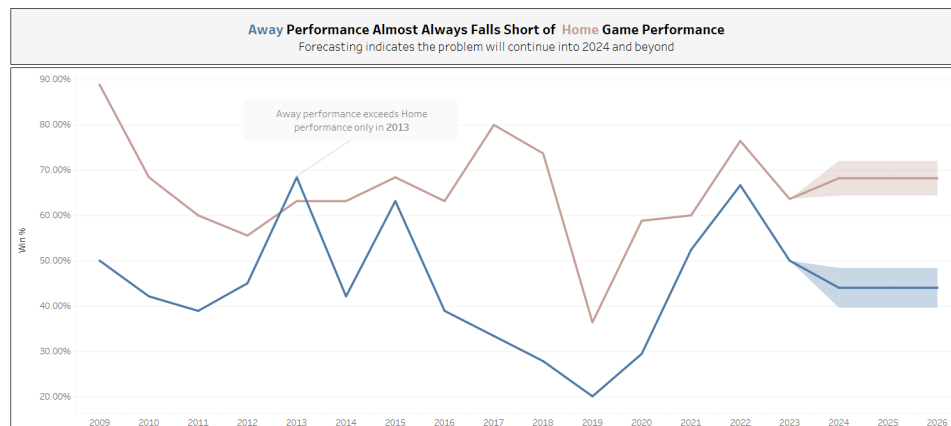


[Project Link](#)

Executive Summary



This analysis identifies areas for Arsenal to improve consistency across venues, aiming to secure their position as the top team in the English Premier League. Insights are tailored for Arsenal's management and coaches, focusing on strategies to achieve more cups, prize money, and increased league prominence, ultimately boosting revenue.

- **Slide 1:** Highlights Arsenal's total wins compared to other top-performing teams, with charts showing total wins and win breakdowns by venue. Identifies competitor weaknesses as opportunities for Arsenal to gain a competitive edge.
- **Slide 2:** Displays Arsenal's win percentage across venues overtime with forecasts for 2024 and beyond, revealing inconsistencies by venue, identifying areas for improvement.
- **Slide 3:** Showcases inconsistencies in Arsenal's offensive performance, highlighting decreased average goals scored and goal conversion rates in away games.
- **Slide 4:** Showcases inconsistencies in Arsenal's defensive performance, highlighting increased average goals allowed and decreased goal prevention rate at away games.
- **Slide 5:** Analyzes correlation between fouls, foul severity, and performance. Includes graphs of win rate, fouls, red cards, and yellow cards by venue to show decreased away game performance can be attributed to increased fouls at away games.

Arsenal struggles with lower performance, more fouls, and greater foul severity in away games. Solutions include enhancing player mental resilience and employing de-escalation strategies to navigate hostile environments effectively.

Basic Info

Project Title: Arsenal Home Vs. Away Performance

Team members: Greg Fagan & Natalie Brown

Data

The dataset we used is from datahub.io, a webpage containing game statistics by season in the English Premier League, we used data from seasons 2009 to 2024.

Description: The dataset is from an API and focuses on soccer team performance. It contains 27 unique attributes, with 21 numerical columns showing metrics such as goals scored and conceded, wins, losses, draws, shots, fouls, yellow and red cards, and halftime performance statistics. It is a medium-scale, structured dataset with moderate cardinality, good for statistical analysis and predictive modeling in sports analytics.

During preprocessing, the original dataset was separated by year, so we merged the data, and we reshaped it from wide format to long format to create a more manageable structure. the year 2019 was removed due to data gaps caused by Covid, which skewed the results. The cleaning was done using Python (with libraries like pandas) and Excel.

Raw Data Source:

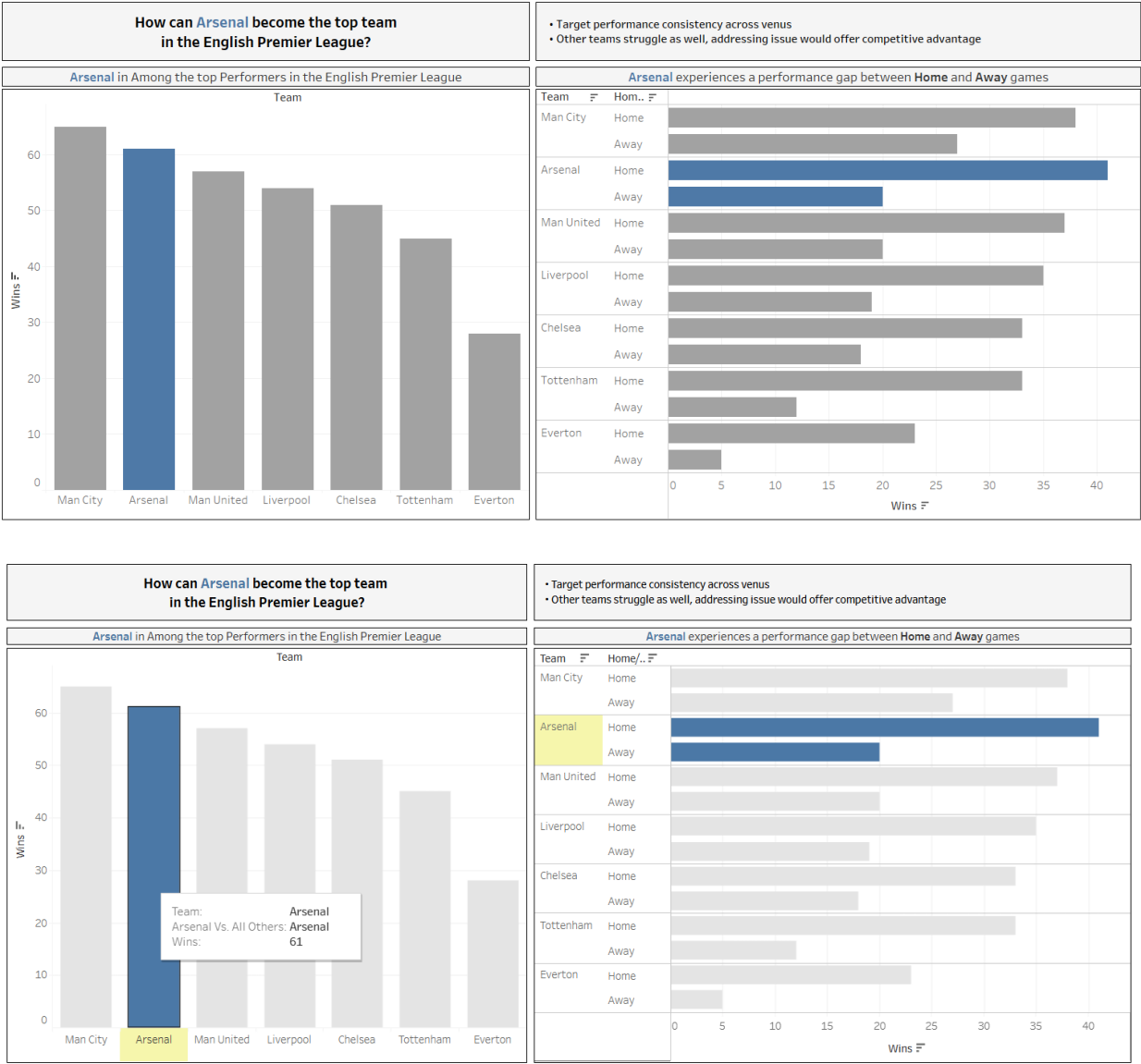
datahub.io

Cleaned Data:

[prem_league_long_09_24.xlsx](#)

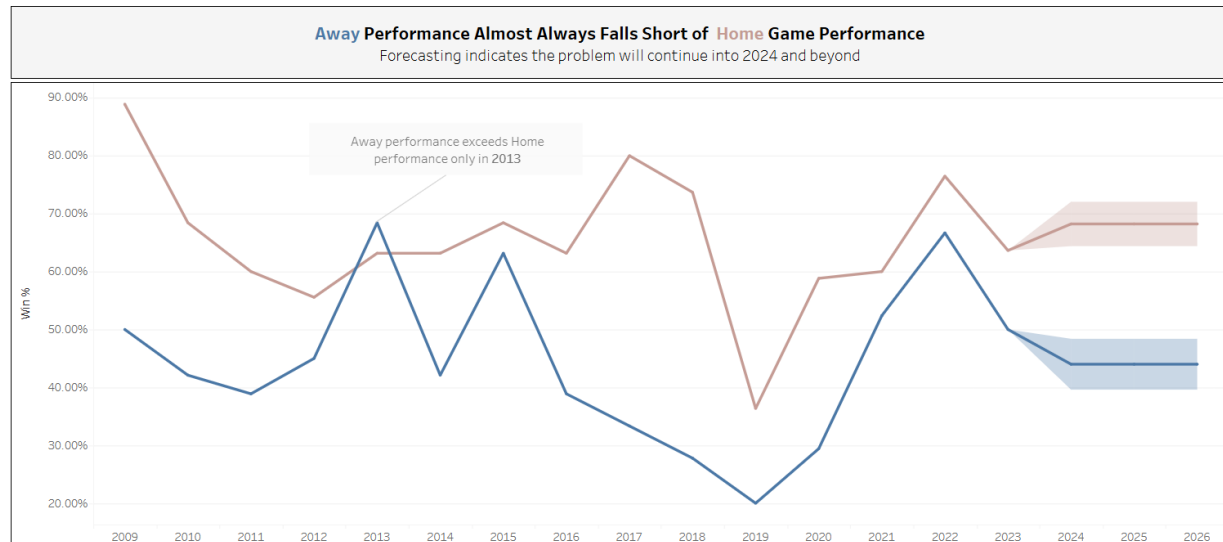
The raw data has a separate csv file for each season, so during preprocessing we combined each season into a single excel file in Python with pandas. Thereafter we transformed the data from wide to long format.

Visualizations

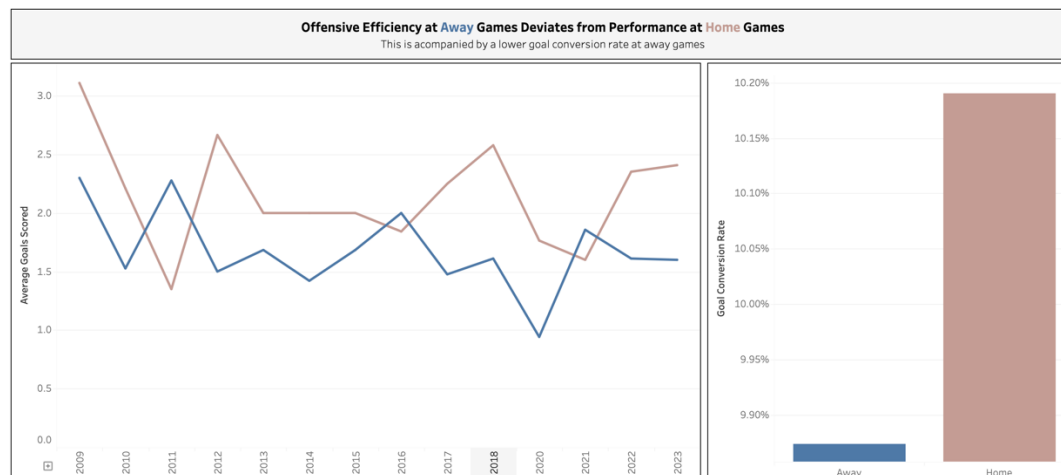
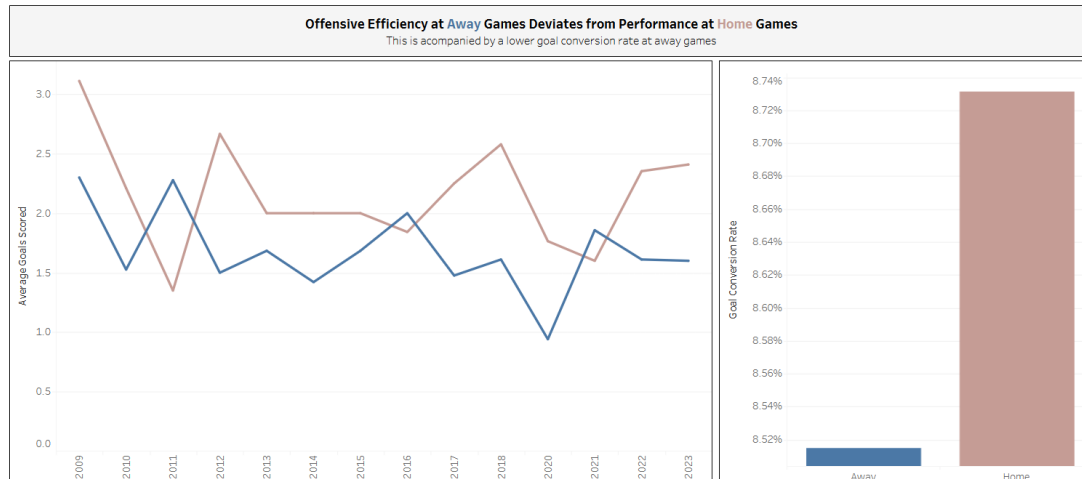


The views in this visualization allow users to look into Arsenal's performance as a top contender in the league, showing the performance gap between home and away games. The visual encoding choices are designed for conciseness: bold font draws attention to key points, making important information easy to spot, while a hue color palette highlights Arsenal's performance in blue, with contextual data displayed in faded grey to put emphasis on the primary data. Bar graphs are used to compare categorical data against quantitative measures, providing a clear and impactful representation of performance. The graph on the right displays the number of wins on the y-axis and teams on the x-axis, while the graph on the left places wins on the x-axis and teams on the y-axis.

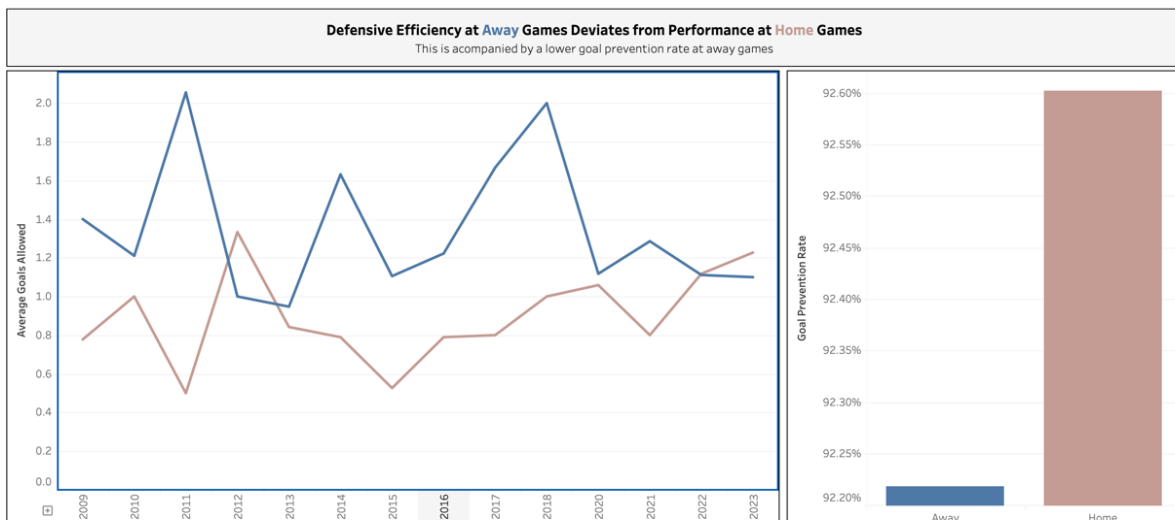
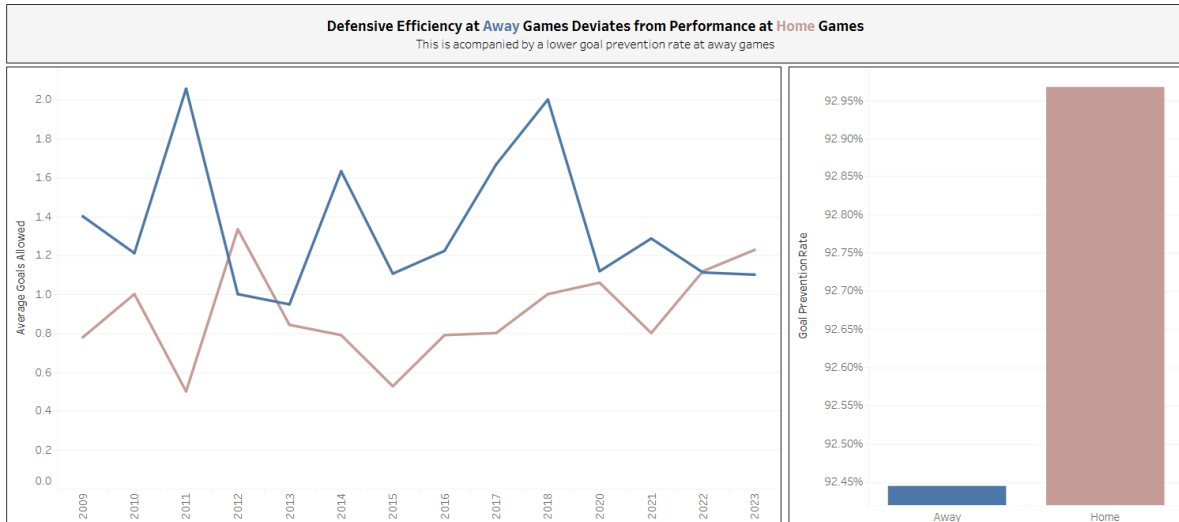
axis, with an additional x-axis differentiating home and away games. These views are linked interactively—when users hover over a team on the left graph, the corresponding stats for that team are highlighted on the right graph, creating a connection between the two visualizations.



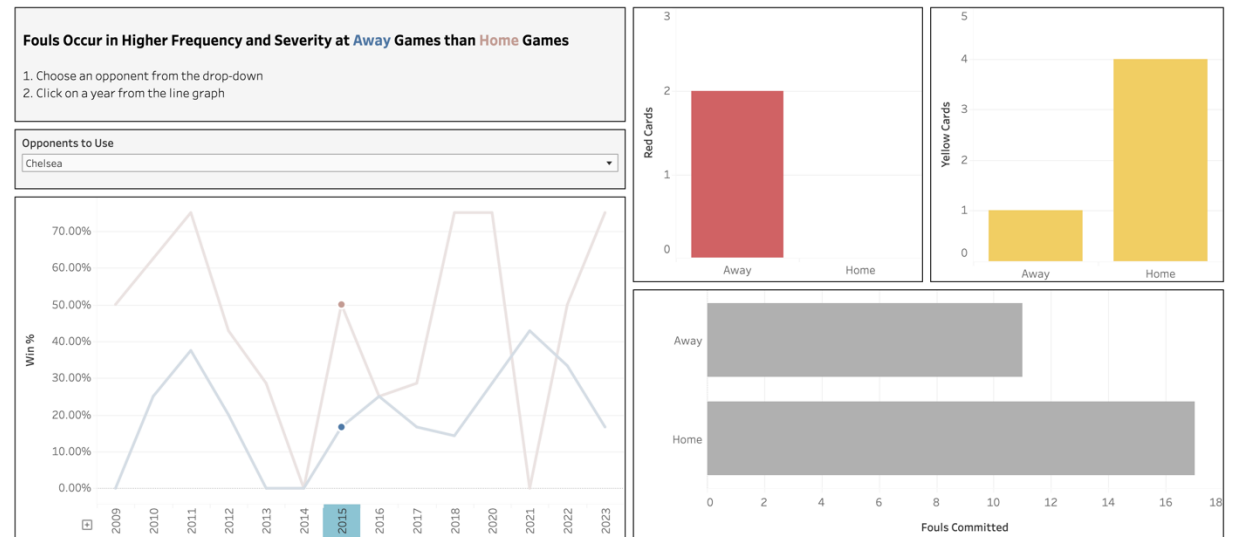
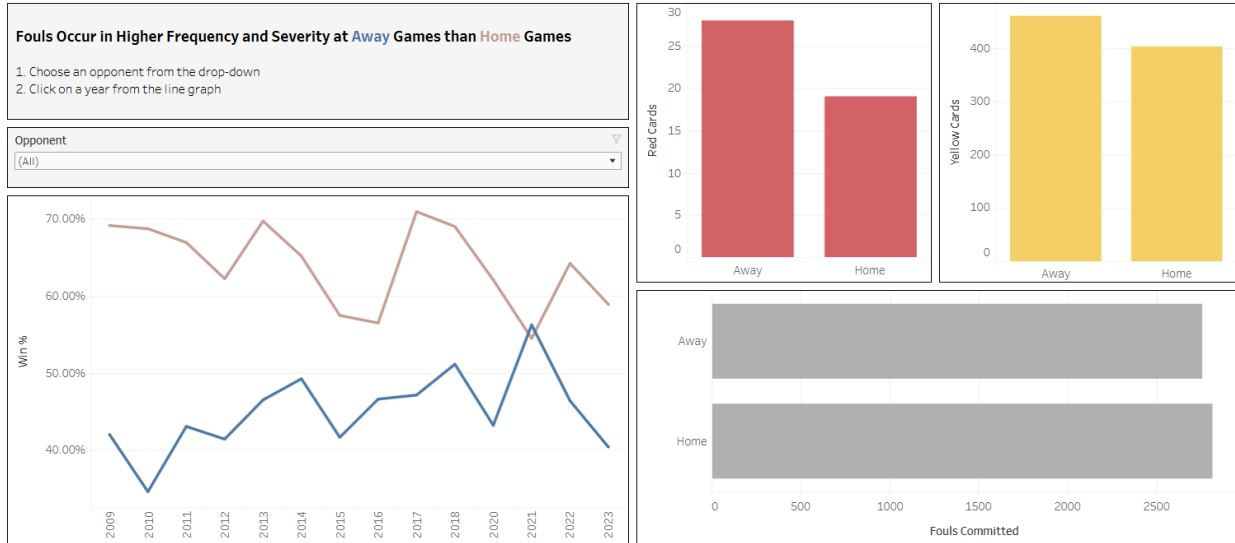
This view illustrates how Arsenal's away performance consistently falls short of their home game performance. The title is color-coded to match the home and away lines, with bold text used to emphasize the key message. Grey scale is applied to contextual information to keep the focus on the main comparison. The visualization uses lines and confidence intervals as marks, with the x-axis representing years and the y-axis showing win percentage. This visualization serves to demonstrate Arsenal's performance over the past 15 years, showing the gap between their home and away results and reinforcing the point that their away performance is weaker. A line graph was chosen for its ability to clearly show trends over time, while the confidence intervals provide valuable insight into the variability and forecasts of the data.



These views allow users to visualize Arsenal's offensive efficiency in away games and how it compares to performance in home games. On the left, a line graph is used to track goal conversion rate over time, chosen for its ability to present quantitative data across periods. Key terms are bold to emphasize important details, and a color palette of blue for home games and red for away games distinguishes the two. Grey contextual information is used to minimize distractions, and legends are replaced with direct labels to reduce visual clutter. A consistent color scheme links the titles and visuals for coherence. On the right, a bar graph is used to compare average goals scored over the years, with the y-axis representing goals scored and the x-axis showing years. The channels on the left represent goal conversion rate on the y-axis and home vs. away on the x-axis. Views are linked interactively: when users hover over a year on the bar graph on the right, the line graph on the left will highlight the specific home and away goal conversion rate gaps for that year.



These views allow users to visualize Arsenal's defensive efficiency in away games and how it compares to performance in home games. On the left, a line graph is used to track goal prevention rate over time, chosen for its ability to clearly present quantitative data across different periods. Key terms are in bold to show the main message, and a color palette of blue for home games and red for away games helps distinguish the two. Grey contextual data reduces distractions, and legends are replaced with direct labels to minimize visual clutter. A consistent color scheme links the titles and visuals for clarity. On the right, a bar graph is used to compare the average goals allowed by Arsenal each year, with the y-axis representing goals allowed and the x-axis representing years. The channels on the left display goal prevention rate on the y-axis and home vs. away on the x-axis. The views are linked interactively: when users hover over a year on the bar graph on the right, the line graph on the left will highlight the specific home and away goal prevention rate gaps for that year.



This view allows users to visualize how Arsenal's fouls occur more frequently and severely in away games compared to home games. The visual encoding uses color to distinguish between "away" and "home," with corresponding-colored lines in the graph. Grey is used to highlight the most important information, and bold text ensures the main message is communicated clearly. Red represents red cards, yellow represents yellow cards, and grey is used for fouls committed. The marks consist of lines on the left graph and bars on the right graphs. The channels on the right include year on the x-axis and win percentage on the y-axis, while the top right graph has home and away on the x-axis and number of cards on the y-axis. The bottom right graph shows fouls committed on the x-axis and home/away on the y-axis. The views are linked by a dropdown menu, allowing users to filter each graph by team. Additionally,

hovering over a year on the left graph will display specific stats for that year on the corresponding graphs to the right.

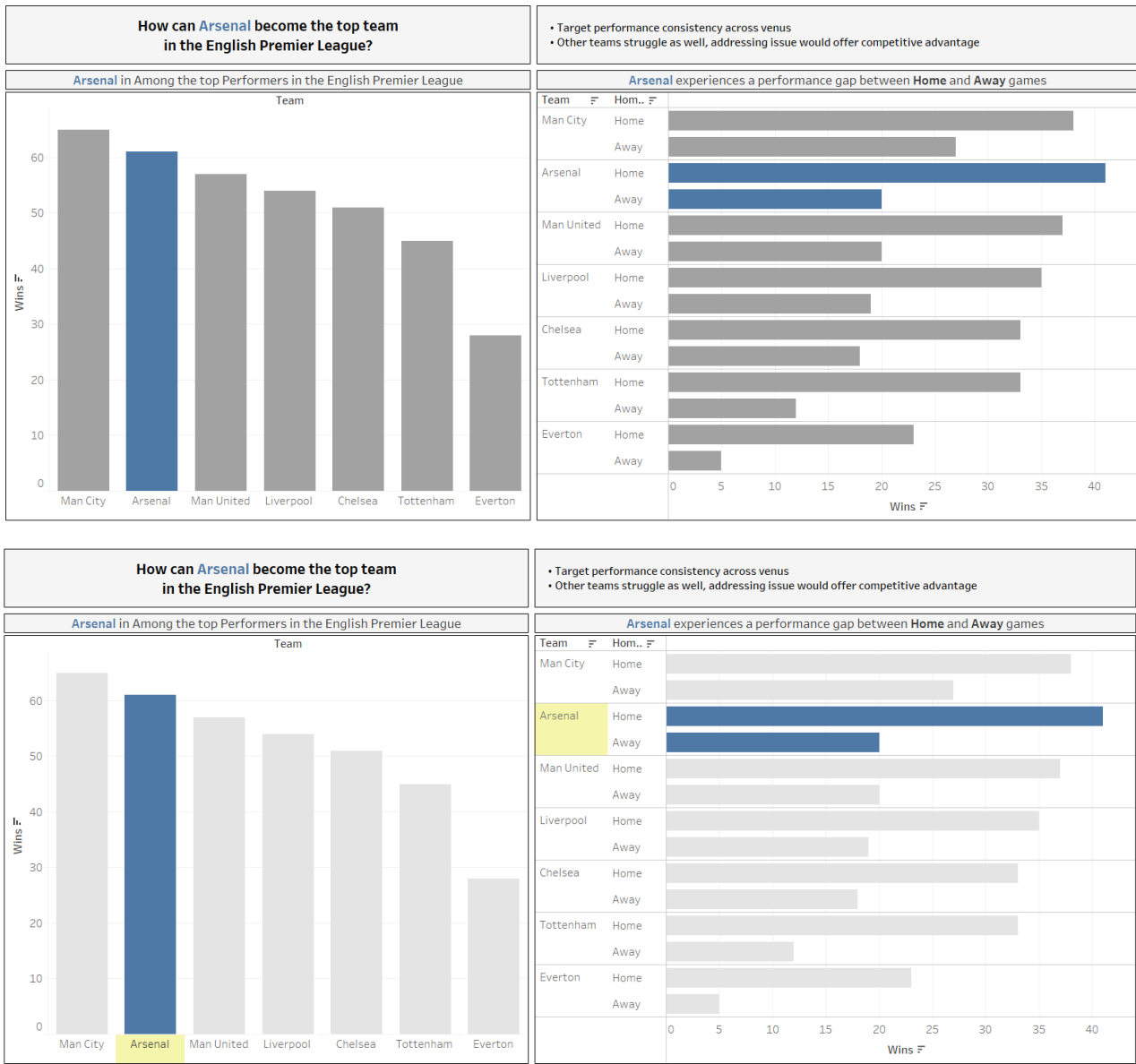
Usage Scenario

Audience: Arsenal management and coaches

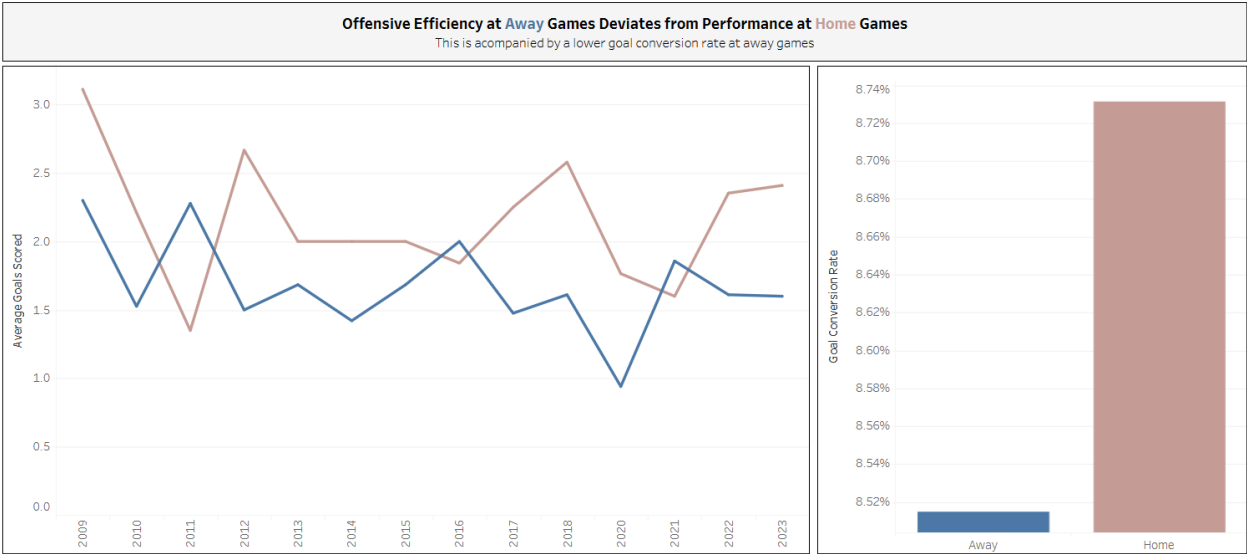
Why They Care: Analysis offers essential insights for Arsenal's leadership to improve performance consistency across venues. Enhancing these areas can help secure the top spot in the English Premier League, win more cups, increase prize money, and boost revenue through heightened league prominence

Action Needed: Invest in programs to enhance players' mental resilience, ensuring they remain unaffected by varying stadium conditions.

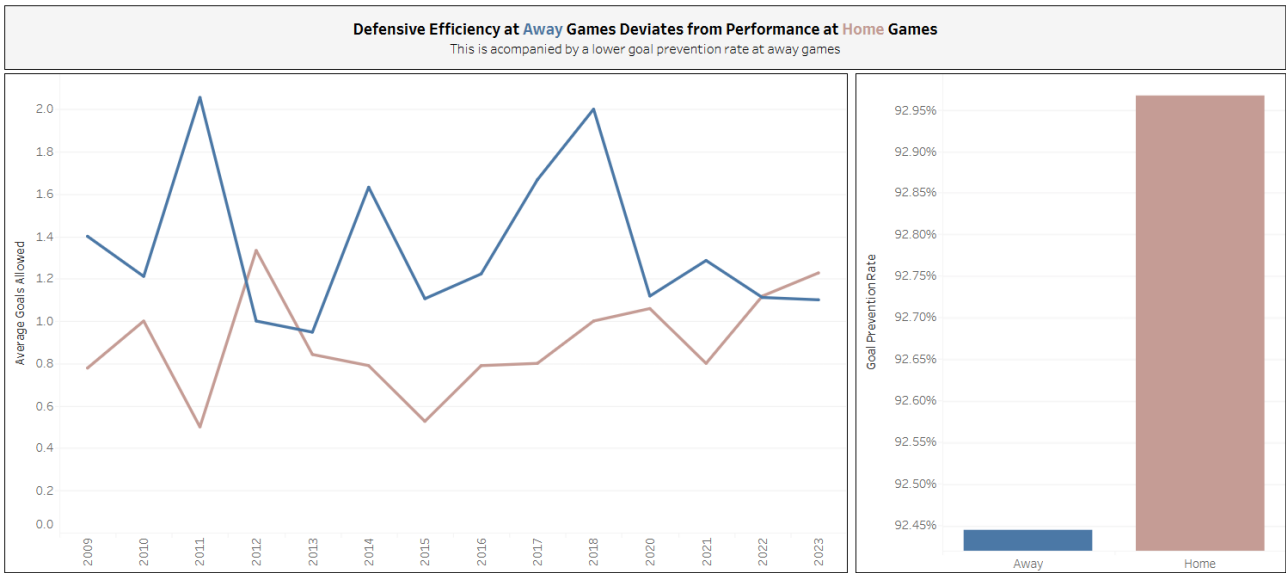
The first slide demonstrates Arsenal's performance compared to top teams and across different venues. The first graph shows Arsenal falling short of the top spot, while the second graph indicates all teams experience decreased performance across venues, highlighting a potential area for Arsenal to outshine the competition. When hovering over a specific team on one graph, the team will be highlighted on the other graph for cohesiveness. We use the highlight focus to first focus on Arsenal and then on the other teams.



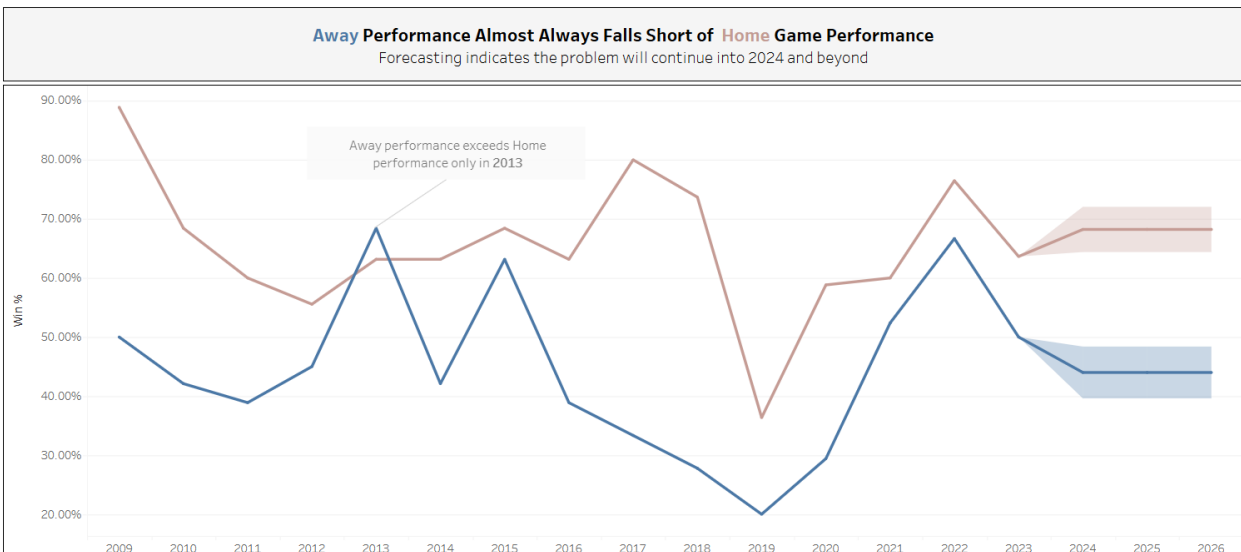
The second slide shows why Arsenal underperforms at away game by analyzing Arsenal's offensive performance across venues, showing average goals over time and goal conversion rates, while focusing on the performance gap.



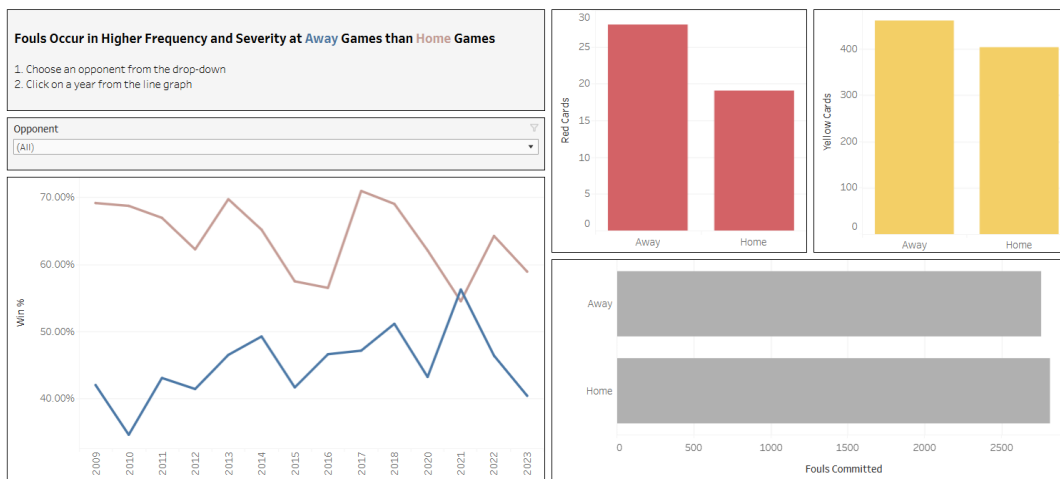
The third slide mirrors the third but focuses on Arsenal's defensive performance gaps, displaying average goals conceded and goal prevention rates.



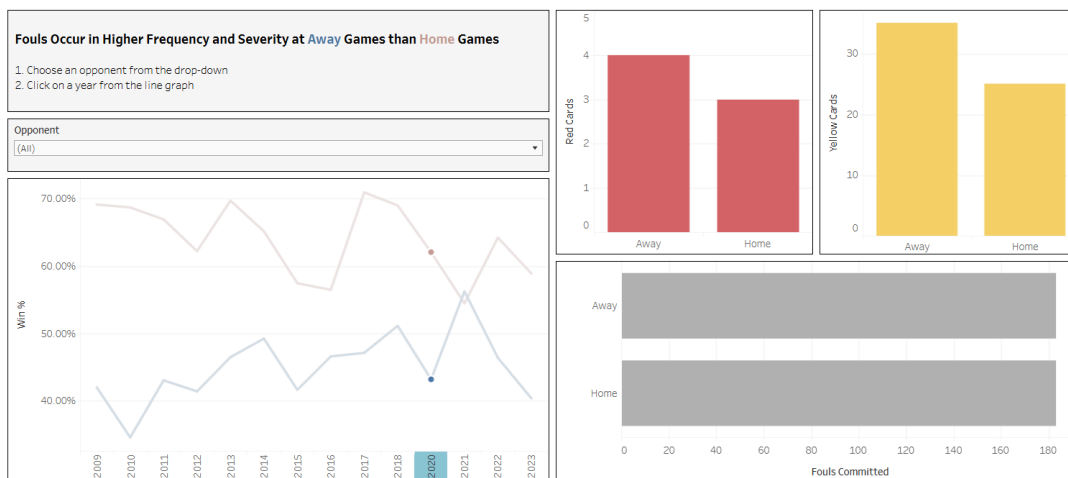
The fourth slide shows how offensive and defensive performance translate to overall performance. Showing win percentage across venues over time highlights away win percentage almost always falls short of home win percentage. Forecasting also indicates that this problem is likely to continue into the future, warranting concern from leadership.



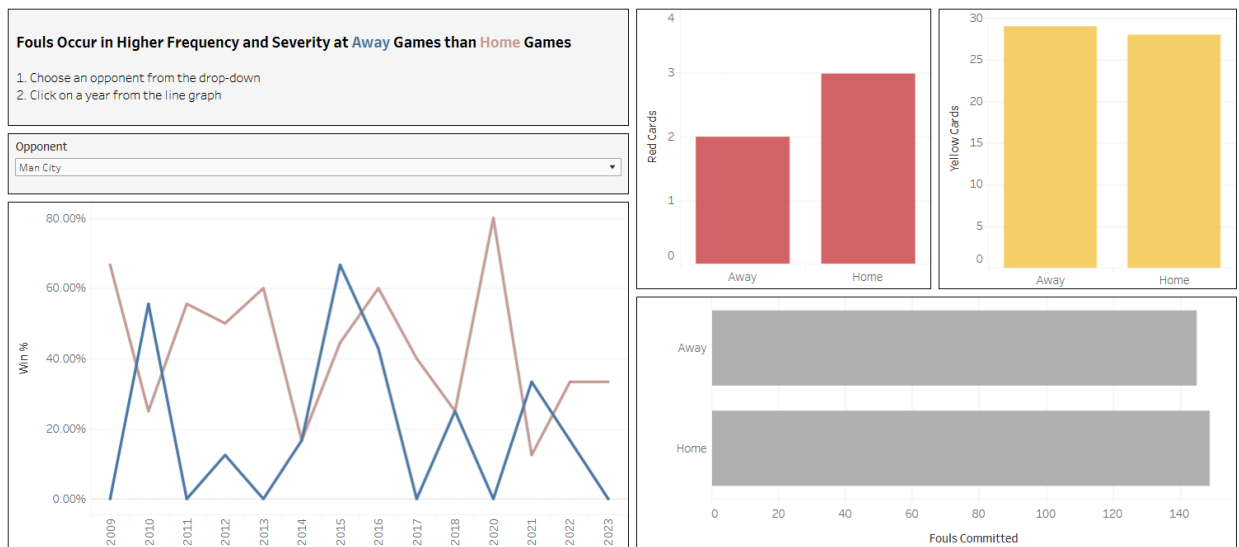
The last slide displays Arsenal's win percentage over the years, accompanied by graphs displaying total fouls by venue, as well yellow cards and red cards.



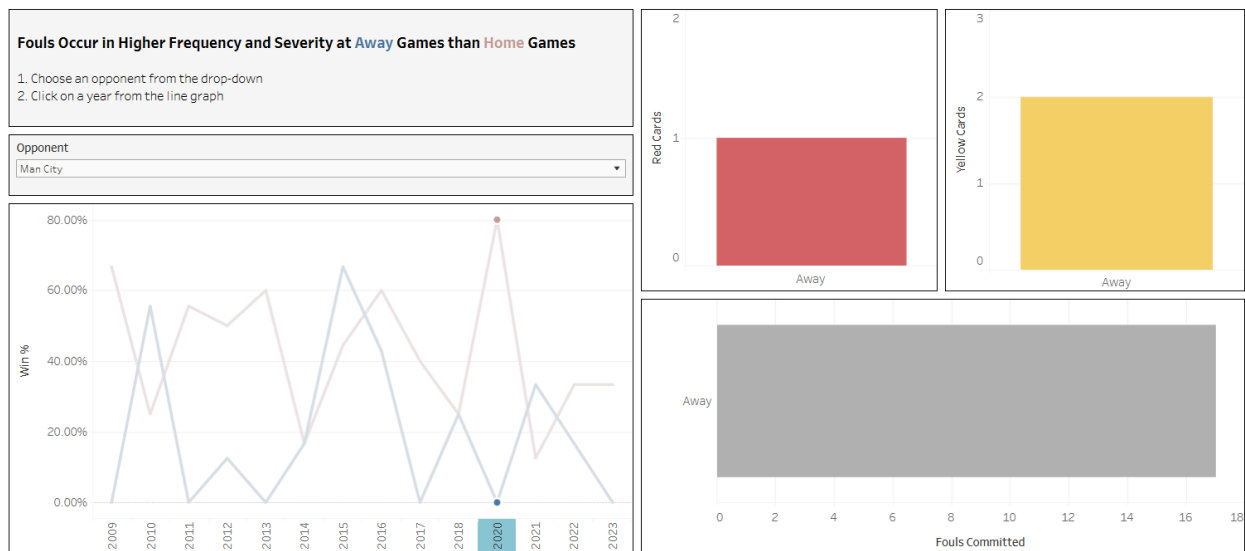
Graphs indicate more severe fouls occur at away games, and when you drill down on year, we are able to see the trend continues in 2020, a year that experienced a lower away game win percentage



We can also filter to show statistics for when our opponent is Manchester City, since they are our top competition.



Then we can filter to a year that has low away performance to see how increased fouls translate to poor performance.



These graphs highlight inconsistencies in performance across venues, indicating a need for programs to enhance players' mental resilience, so they can remain immune to the hostile away stadiums they compete in.

Reflection

- What was the most enjoyable part of the project? And the least enjoyable?

The most enjoyable part of the project was the exploratory data analysis. Being able to visualize the data and understand the trends that you couldn't see before, ignites your curiosity and creativity.

- Describe how your project has developed from your initial proposal, through your first submission, to your final product.

Our first proposal focused on the league as a whole, trying to identify trends and insights that could be applied to all teams. As we progressed, we realized it would be more valuable to focus on a single team, Arsenal, and provide detailed insights tailored to their performance. While we still considered some comparisons with their competition, the main idea went to showing detailed insights specific to Arsenal's performance.

- How have your visualization goals changed?

Our original visualization goals were to focus on team-level performance metrics across the league. Using more categorical features in our graphs to represent each team, we couldn't be as detailed or flexible with our numeric graphs. But when we decided to focus our visualizations specific to Arsenal, it allowed us room to create detailed interactive dashboards centered around Arsenal's numeric performance rather than a bland general league comparison.

- How have your technical goals changed?

At first, our technical goals focused on cleaning data and building visualizations for a league-wide approach. As we switched to focusing on one team, our goals evolved to restructure the data to fit this new focus, and creating in-depth dashboards and specific interactive visuals that would clearly communicate our findings from the EDA about Arsenal.

- Was there anything you wanted to implement that you couldn't figure out how to do? If so, then what workarounds did you employ, or did you abandon your original idea?

We wanted to create more visualizations that included player-level performance alongside team-level stats, but we couldn't figure out how to join the data due to the complexity of combining multiple

datasets, while making sure the data was still accurate. Because of time constraints, we abandoned the idea.

- What would you do differently if you were to make the project again from scratch (or any other interactive visualization)?

If we were starting from scratch, we would focus more on including player-level data, because this would provide deeper insights into Arsenal's performance. Player personnel is a big factor in team success, and providing this data about them would have strengthened the presentation by showing individual contributions and their impact on overall team performance.

Project Management & Team Assessment

Task Description	Estimated Hours	Actual Hours	Completion Date	Assigned Team Members & Responsibilities
Project Planning	1	1	11/17/24	Natalie & Greg - Defined project proposal, goals, and deliverables.
Data Collection	1	2	11/20/24	Greg - Gathered dataset from Kaggle
Data Cleaning	2	5	12/1/24	Natalie - Preprocessed and cleaned the data using Python and Excel.
Data Exploration & Analysis	2	3	12/5/24	Greg - Performed exploratory data analysis and created insights using Python and tableau.
Data Restructuring	2	2	12/7/24	Natalie – Restructured data to fit new goal.
Visualization Development	3	5	12/9/24	Natalie & Greg- Designed and created interactive Tableau dashboards. Allocating certain graphs to each other.
Report Writing	2	3	12/10/24	Greg & Natalie- Worked on report. We split up report by sections.
Final Review	1	1	12/11/24	Natalie & Greg - Reviewed dashboards and story accuracy and presentability. Practiced the presentation too.
Total	14	22		