Valorant Champions 2024 Seoul

Comprehensive Data Analysis Report

Author: Piyush Kumar
Domain: Esports Analytics (Competitive FPS – Valorant)
Tools: Python, Pandas, Numpy, Matplotlib, Seaborn

Table of Contents

1. Executive Summary

This project analyzes data from the **Valorant Champions 2024 (VCT Seoul)** tournament to identify the key factors that contribute to winning at the highest level of professional play. Using match, player, agent, map, and economy datasets, the analysis focuses on player performance, strategic agent and map selection, and economic decision-making rather than raw kill counts alone. The findings show that metrics such as **KAST% and ADR** are stronger indicators of impact, effective **agent–map synergies** outperform meta popularity, and efficient economy management—particularly in Thrifty rounds—provides a measurable competitive advantage. This project demonstrates strong data engineering, exploratory analysis, and insight-driven storytelling skills.

2. Project Objective & Business Context

Objective

The primary objective of this project was to build a structured and repeatable analytical framework capable of explaining why certain teams consistently win at the highest level of competitive Valorant. Rather than relying on surface-level indicators such as total kills or highlight performances, the analysis was designed to uncover underlying performance drivers across three dimensions:

- Identifying the key statistical differences between winning and losing teams
- Determining which player-level metrics (e.g., KAST%, ADR) best represent true in-round impact
- Evaluating how strategic decisions, including agent composition and economy management, influence match outcomes
- This approach ensures that insights are grounded in efficiency, consistency, and decision quality rather than isolated events.

Business Context & Relevance

- From a business and analytics perspective, this project closely mirrors real-world data analysis challenges:
- Data is fragmented across multiple sources, requiring careful integration and validation
- Raw or headline metrics can be misleading without proper context and normalization
- Stakeholders require actionable insights that support strategic decisions, not just descriptive dashboards

By transforming complex tournament data into clear, decision-oriented insights, this project demonstrates how data analytics can be used to support strategy evaluation, performance

optimization, and predictive decision-making—skills directly transferable to professional Data Analyst and Data Science roles.

3. Data Sources & Project Structure

Data Sources

The dataset is composed of multiple structured CSV files, each representing a specific analytical layer of the Valorant Champions 2024 (VCT Seoul) tournament. This modular structure enables flexible joins and supports both high-level and granular analysis across matches, maps, players, agents, and economy.

Core Reference & Metadata Files

- **`event_info.csv`**
  Contains general information about the tournament, including event name, location, and format. This file serves as the top-level metadata reference for the analysis.

- **`columns_description.csv`**
  Provides detailed descriptions of all columns across the dataset. This file improves dataset interpretability, supports maintainability, and ensures clarity when working across multiple analytical notebooks.

Match & Map-Level Data

- **`matches.csv`**
  Stores match-level information such as match identifiers, participating teams, and match outcomes. This file acts as a primary linkage point for integrating player, map, and economy data.

- **`detailed_matches_maps.csv`**
  Contains map-specific data for each match, including map names, results, and team performance. This dataset enables map-wise performance analysis and terrain-specific trend identification.

- **`maps_stats.csv`**
  Aggregated statistics for each map, including win percentages for attack and defense sides. This file supports strategic analysis of map balance and side advantages.

Player-Level Performance Data

- **`player_stats.csv`**
  Provides overall player statistics for the tournament, offering a high-level view of individual performance and enabling leaderboard creation and outlier detection.

- **`detailed_matches_player_stats.csv`**
  Contains granular, match-level player statistics, including ACS, ADR, KAST%, and K/D. This file is central to evaluating player efficiency and consistency across different maps and matches.

- **`performance_data.csv`**
  Includes advanced performance indicators such as multi-kill rounds and clutch scenarios. This dataset enables deeper analysis of high-impact moments beyond standard metrics.

Agent & Strategic Data

- **`agents_stats.csv`**
  Contains statistics for each agent, including pick rates and win rates. This file supports analysis of agent popularity, effectiveness, and agent–map synergies.

Economy Data

- **`economy_data.csv`**
  Captures team economy information at the map and round level, including credits spent and buy types. This dataset is critical for analyzing the relationship between economic decision-making and round win probabilities, including Thrifty round impact.

Analytical Value

The separation of data into specialized CSV files reflects a **normalized data design**, similar to real-world production systems. This structure required deliberate data engineering effort to reconstruct meaningful analytical views, reinforcing skills in:

- Multi-table joins using primary and foreign keys
- Data validation and schema understanding
- Building scalable, reusable analytical pipelines

Data Collection Methodology

The data was collected through a custom-built web scraping pipeline developed specifically for this project. Match, player, and economy data were scraped from the vlr.gg website, one of the most authoritative sources for professional Valorant esports statistics.

The scraper was implemented as a separate data engineering project, demonstrating ownership of the full data lifecycle:

- **Scraper Repository:** https://github.com/Piyush86kumar/Valorant_vlr.gg_scrapper
- **Output Format:** Structured CSV files optimized for downstream analysis
- **Design Focus:** Reliability, reusability, and schema consistency across tournaments

Once collected, the data was cleaned, validated, and versioned before being uploaded to Kaggle for community use, collaboration, and reproducible analysis.

Project Structure

```
VCT 2024/
│
├── VCT_2024_seoul_dataset/                    # Dataset storage
│   ├── agents_stats.csv                  # Agent utilization and win rates
│   ├── columns_description.csv           # Metadata and column definitions
│   ├── detailed_matches_maps.csv         # Specific map data per match
│   ├── detailed_matches_overview_processed.csv # Cleaned/Transformed match overview
│   ├── detailed_matches_overview.csv     # Raw match overview data
│   ├── detailed_matches_player_stats.csv # Match-by-match player performance
│   ├── economy_data.csv                  # Team economy and buy-type stats
│   ├── event_info.csv                    # VCT tournament general info
│   ├── maps_stats.csv                    # Map win % (Attack vs. Defense)
│   ├── matches.csv                       # List of all matches played
│   ├── performance_data.csv              # Clutches and multi-kill data
│   └── player_stats.csv                  # Aggregated player statistics
│
├── Agents_stats.ipynb                    # Analysis of agents_stats.csv
├── Detailed_matches_maps.ipynb           # Analysis of maps and match flow
├── Detailed_player_stats.ipynb           # Analysis of performance_data.csv
├── Economy_stats.ipynb                   # Analysis of economy_data.csv
├── Player_stats.ipynb                    # Analysis of player_stats.csv
├── README.md                             # Project documentation
└── requirements.txt                      # Necessary Python libraries
```

4. Data Engineering & Preparation

Data Integration

Merged 5+ CSV files using keys such as:

match_id

player_id

Map_id

Created unified datasets at:

Player–map level

Team–map level

Data Cleaning

Converted percentage strings (e.g., "78%") into numeric values

Handled missing map and agent fields

Normalized player stats to ensure fair comparison across matches

Feature Engineering

Derived efficiency-focused metrics:

ACS consistency

ADR stability

Prepared data for outlier and distribution analysis

5. Exploratory Data Analysis Overview

Initial EDA focused on:

Metric distributions

Missing value patterns

Variance across players and maps

This step ensured statistical validity before deeper insights were extracted.

6. Pillar 1: Player Performance & Efficiency

(Detailed_player_stats.ipynb, Player_stats.ipynb)

Metrics Used

- ACS (Average Combat Score)
- ADR (Average Damage per Round)
- KAST%
- K/D Ratio

Key Findings

- **KAST% is a stronger indicator of winning impact than K/D**
- Elite players distinguish themselves through **consistency across metrics**
- High ACS alone does not guarantee map wins

Distribution analysis revealed **clear statistical outliers**, such as dominant performances on specific maps (e.g., Haven).

---

7. Pillar 2: Agent Meta & Map Strategy

(Agents_stats.ipynb, Detailed_matches_maps.ipynb)

Analysis Focus

- Agent pick rate vs win rate
- Agent effectiveness by map

Key Findings

- Popular agents are not always the most effective

- Strong **agent–map synergies** exist
- Top teams prioritize **situational compositions** over meta imitation

This highlights the importance of **strategic preparation**.

---

8. Pillar 3: Economy & Round Win Impact

(Economy_stats.ipynb)

Analysis Focus

- Team credits vs round wins
- Thrifty rounds and momentum

Key Findings

- Higher spending shows diminishing returns
- Thrifty wins significantly improve future round win probability
- Efficient teams consistently win with lower average spend

This demonstrates **resource efficiency as a competitive advantage**.

---

9. Visualization Strategy

The project uses purpose-driven visualizations:

- 2×2 subplot grids for metric comparison
- Horizontal bar charts with value labels
- Heatmaps for agent-map effectiveness

- Distribution plots for outlier detection

Each visualization was created to **answer a specific analytical question**, not for aesthetics alone.

---

10. Key Insights & Business Value

Strategic Insights

- Efficiency beats aggression
- Economy discipline amplifies momentum
- Meta understanding must be contextual

Business Value

This analysis framework can support:

- Team strategy planning
- Player evaluation
- Predictive match modeling

---

11. Technical Toolkit

Languages & Libraries

- Python
- Pandas
- NumPy

- Matplotlib
- Seaborn

Techniques

- Exploratory Data Analysis (EDA)
- Feature Engineering
- Metric Normalization
- Outlier Detection
- Analytical Storytelling

---

12. Limitations

- No round-by-round positional data
- Limited temporal modeling
- External factors (comms, pressure) not captured

---

13. Future Scope

- Build a **Win Probability Prediction Model**
- Use agent composition + economy features
- Apply Logistic Regression / Tree-based models
- Add momentum and time-series features