

Moneyball – Steps to Execute

Data Info

All data has been extracted from UMTerps Softball website. All the way from 1995-2023, each year's schedule has been extracted, cleaned, pre-processed and analyzed.

Data Source: <https://umterps.com/sports/softball/schedule>

Data Collection: The data is collected using a web scraping tool on the 'umterps' website.

Extraction Method: Web Scraper Chrome Extension to extract by HTML selectors.

Data Format: The data is downloaded and stored in .xlsx format.

There is a single derived column in the Moneyball.Game table which takes input from gmeTime and stores it in =gmeTimeOfDay using the following convention:

If the time is more than 00:01 but less than 12:00, it returns "Morning."

If the time is between 12:00 and 17:00, it returns "Afternoon."

If the time is between 17:00 and 20:00, it returns "Evening."

If the time is between 20:00 and 00:01, it returns "Night."

Run Attached SQL Files in this order:

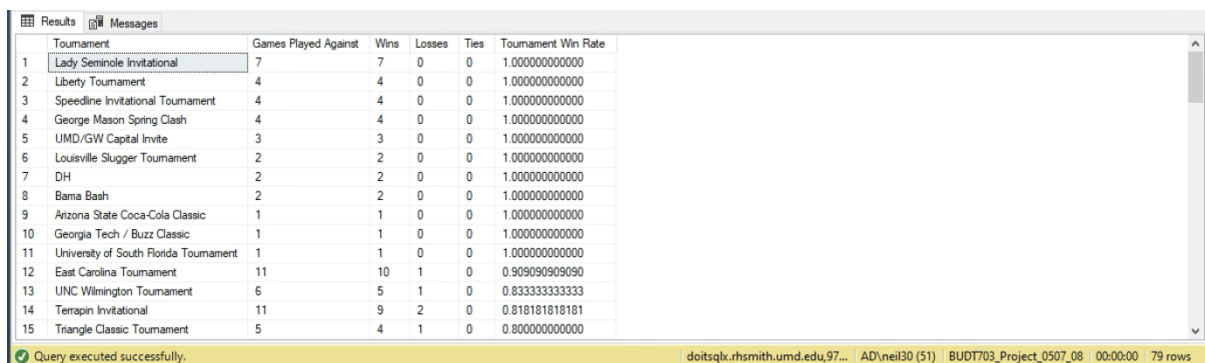
1. Project 0507_08_Create.sql

As CTEs don't show up in the Tableau file, we replicated the select statements for win/losing streaks in the form of CREATE VIEW statements in this file. The following CREATE VIEWS do not have any bearing on the Select code, however they help in replicating outputs in Tableau: LongestWinStreaks, LongestLossStreaks, ConsecutiveWinGroups, ConsecutiveLossGroups, RankedGames, Moneyball_Game_View

2. Project 0507_08_Insert.sql
3. Project 0507_08_Select.sql

Outputs should match:

- 1) What tournaments did the UMD softball team perform best in?



The screenshot shows a SQL query results window with a table containing tournament performance data. The table has columns for Tournament, Games Played Against, Wins, Losses, Ties, and Tournament Win Rate. The data is sorted by Tournament Win Rate in descending order.

	Tournament	Games Played Against	Wins	Losses	Ties	Tournament Win Rate
1	Lady Seminole Invitational	7	7	0	0	1.000000000000
2	Liberty Tournament	4	4	0	0	1.000000000000
3	Speedline Invitational Tournament	4	4	0	0	1.000000000000
4	George Mason Spring Clash	4	4	0	0	1.000000000000
5	UMD/GW Capital Invite	3	3	0	0	1.000000000000
6	Louisville Slugger Tournament	2	2	0	0	1.000000000000
7	DH	2	2	0	0	1.000000000000
8	Bama Bash	2	2	0	0	1.000000000000
9	Arizona State Coca-Cola Classic	1	1	0	0	1.000000000000
10	Georgia Tech / Buzz Classic	1	1	0	0	1.000000000000
11	University of South Florida Tournament	1	1	0	0	1.000000000000
12	East Carolina Tournament	11	10	1	0	0.909090909090
13	UNC Wilmington Tournament	6	5	1	0	0.833333333333
14	Terapin Invitational	11	9	2	0	0.818181818181
15	Triangle Classic Tournament	5	4	1	0	0.800000000000

Query executed successfully. doitsqlx.rhsmith.umd.edu,97... AD\neil30 (51) BUDT703_Project_0507_08 00:00:00 79 rows

2) What is the win rate of the UMD team with and without home field advantage?

Results

Messages

	Location Type	Games Played	Wins	Losses	Ties	Win Rate
1	Home	652	395	257	0	0.605828220858
2	Away	533	253	280	0	0.474671669793
3	Neutral	265	117	146	2	0.441509433962

Query executed successfully.

doitsqlx.rhsmith.umd.edu,97... AD\neil30 (51) BUDT703_Project_0507_08 00:00:00 3 rows

3) Who are the top 10 weakest opponents against UMD?

Results		Messages					
	Opponent ID	Opponent Team	Games Played Against	Wins	Losses	Ties	Win Rate Against Opponent
1	OPP0167	Delaware State	19	19	0	0	1.000000000000
2	OPP0168	Howard	9	9	0	0	1.000000000000
3	OPP0054	Coppin State	9	9	0	0	1.000000000000
4	OPP0200	MOUNT ST. MARY'S	8	8	0	0	1.000000000000
5	OPP0261	UNC Charlotte	7	7	0	0	1.000000000000
6	OPP0044	Coastal Carolina	7	7	0	0	1.000000000000
7	OPP0088	Rhode Island	6	6	0	0	1.000000000000
8	OPP0194	Connecticut	5	5	0	0	1.000000000000
9	OPP0241	La Salle	4	4	0	0	1.000000000000
10	OPP0082	Nevada	4	4	0	0	1.000000000000
11	OPP0133	Mercer	4	4	0	0	1.000000000000
12	OPP0011	Saint Joseph's	3	3	0	0	1.000000000000
13	OPP0015	Ball State	3	3	0	0	1.000000000000
14	OPP0246	Morgan St	3	3	0	0	1.000000000000
15	OPP0231	Maryland-Eastern Shore	3	3	0	0	1.000000000000

Query executed successfully.

doitsqlx.rhsmith.umd.edu,97... AD\neil30 (51) BUDT703_Project_0507_08 00:00:00 275 rows

4) Who are the top 10 strongest opponents against UMD?

Results

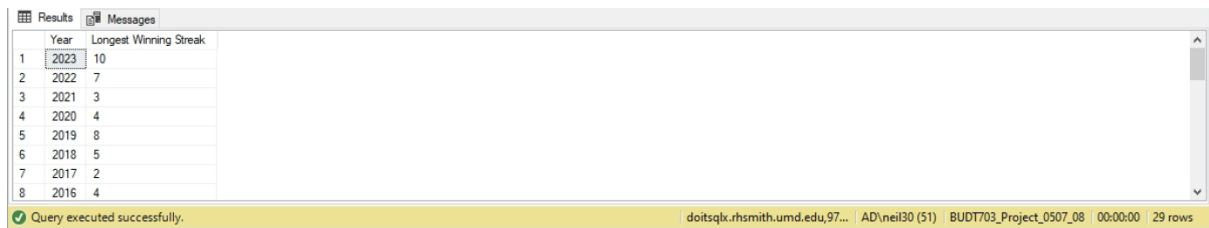
Messages

	Opponent ID	Opponent Team	Games Played Against	Wins	Losses	Ties	Win Rate Against Opponent
1	OPP0147	Notre Dame	8	0	8	0	0.000000000000
2	OPP0101	Louisville	4	0	4	0	0.000000000000
3	OPP0150	Mississippi State	4	0	4	0	0.000000000000
4	OPP0161	UCLA	4	0	4	0	0.000000000000
5	OPP0274	Anne Arundel CC	4	0	4	0	0.000000000000
6	OPP0099	Illinois State	3	0	3	0	0.000000000000
7	OPP0103	Miami OH	3	0	3	0	0.000000000000
8	OPP0107	Tennessee	3	0	3	0	0.000000000000
9	OPP0090	#20 Minnesota	3	0	3	0	0.000000000000
10	OPP0120	Oregon	3	0	3	0	0.000000000000
11	OPP0123	UT Martin	2	0	2	0	0.000000000000
12	OPP0141	Oklahoma	2	0	2	0	0.000000000000
13	OPP0066	Clemson	2	0	2	0	0.000000000000
14	OPP0072	Arizona State	2	0	2	0	0.000000000000
15	OPP0275	Trenton State	2	0	2	0	0.000000000000

Query executed successfully.

doitsqlx.rhsmith.umd.edu,97...AD\neil30 (51)BUDT703_Project_0507_0800:00:00275 rows

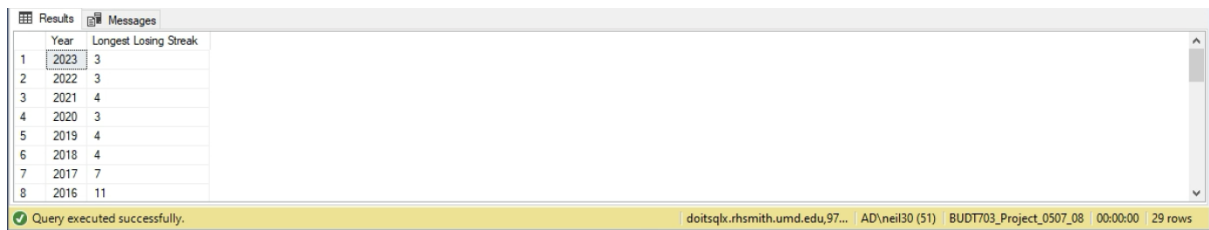
5) What are the longest winning streaks by year?



Query executed successfully. doitsqlx.rhsmith.umd.edu,97... AD\neil30 (51) BUDT703_Project_0507_08 00:00:00 29 rows

	Year	Longest Winning Streak
1	2023	10
2	2022	7
3	2021	3
4	2020	4
5	2019	8
6	2018	5
7	2017	2
8	2016	4

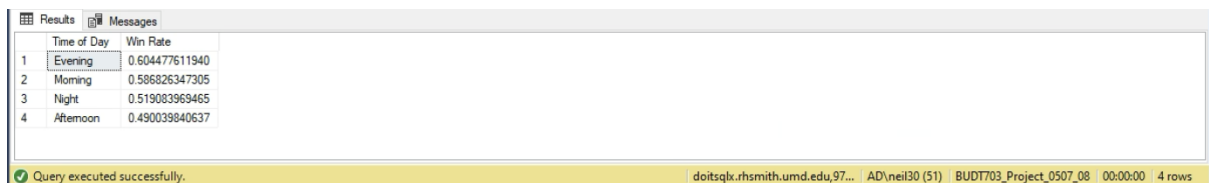
6) What is the longest losing streak for each year?



Query executed successfully. doitsqlx.rhsmith.umd.edu,97... AD\neil30 (51) BUDT703_Project_0507_08 00:00:00 29 rows

	Year	Longest Losing Streak
1	2023	3
2	2022	3
3	2021	4
4	2020	3
5	2019	4
6	2018	4
7	2017	7
8	2016	11

7) What time of day does the UMD softball team perform its best?



Query executed successfully. doitsqlx.rhsmith.umd.edu,97... AD\neil30 (51) BUDT703_Project_0507_08 00:00:00 4 rows

	Time of Day	Win Rate
1	Evening	0.604477611940
2	Morning	0.586826347305
3	Night	0.519083969465
4	Afternoon	0.490039840637

Tableau:

Tournament Performance (Q1)

Data Source:

TournamentWinRate (BUDT703_Project_0507_08)

TournamentWinRate



The CREATE VIEW for TournamentWinRate has all the necessary fields.

Configuration:

Columns

SUM(Win Rate)

Rows

Tournament

Marks

Bar

Color

Size

Label

Detail

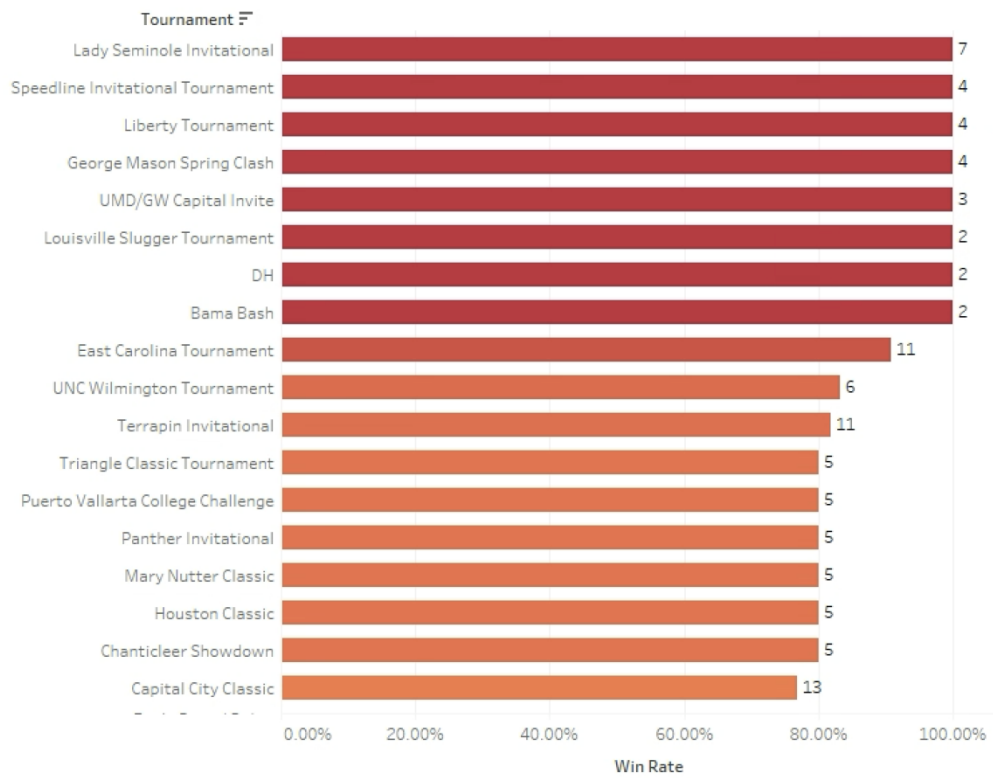
Tooltip

SUM(Win Rate)

SUM(Games Pl..)

Visualization:

What tournaments did the UMD softball team perform best in?



Home Field Advantage (Q2)

Data Source:

LocationWiseScore

The CREATE View for LocationWiseScore contains all the necessary fields.

Configuration:

ColumnsTime Of Day

RowsAGG(Win Rate)

Filters

HOUR(Gme Time)

Marks

Circle

ColorSizeLabelDetailTooltip

AGG(Win Rate)AGG(Win Rate)

Top 10 Weakest/Strongest Opponents (Q3, Q4)

Weakest/Strongest Opponent

Connection

LiveExtract

OpponentWinRate

Moneyball.Opponent

Moneyball.Hold

Moneyball.Game

Moneyball.Tournament

Moneyball.Location

LocationWiseScore

ColumnsSUM(Games Played ..

RowsOpponent Team

Marks

Bar

Color Size Label

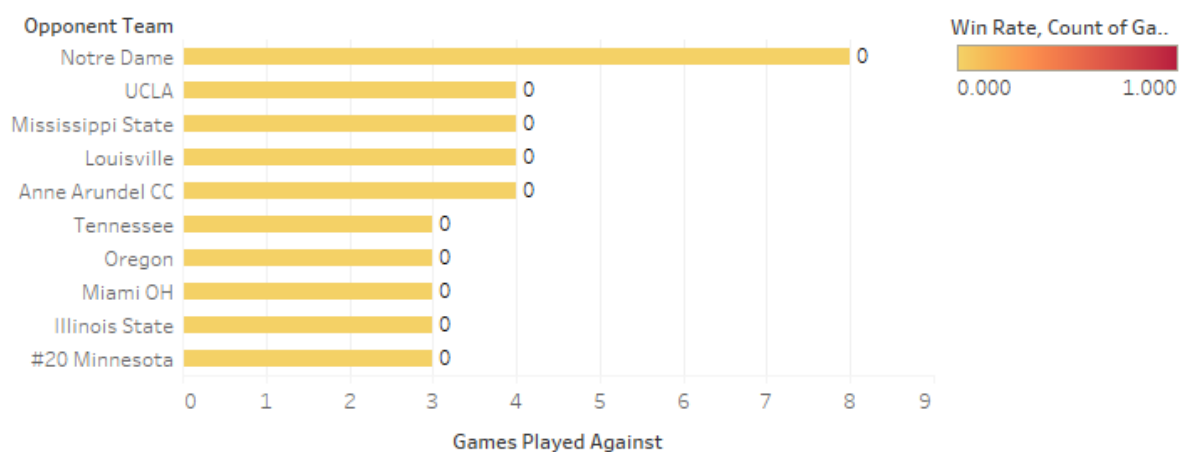
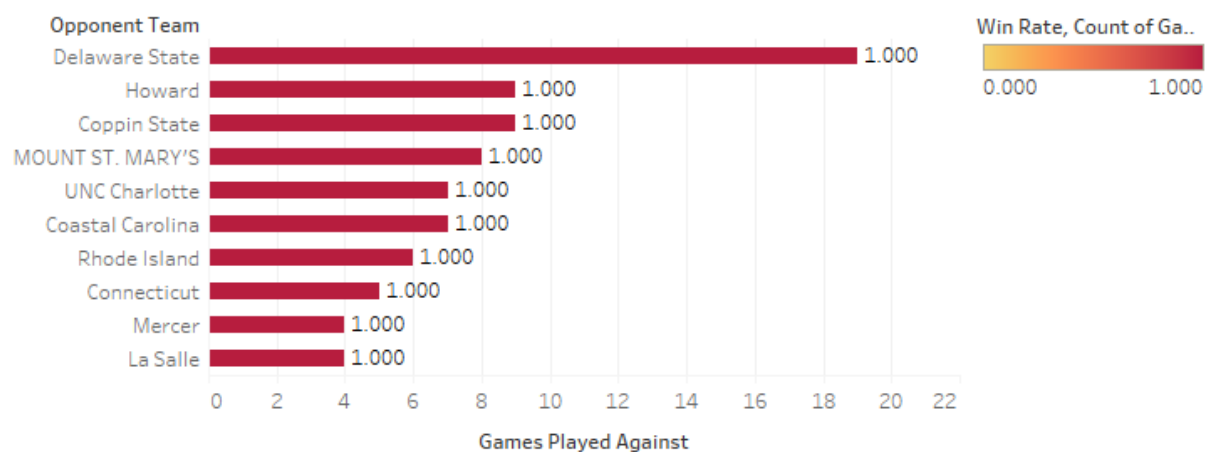
Detail Tooltip

SUM(Win Rate ..)

SUM(Win Rate ..)

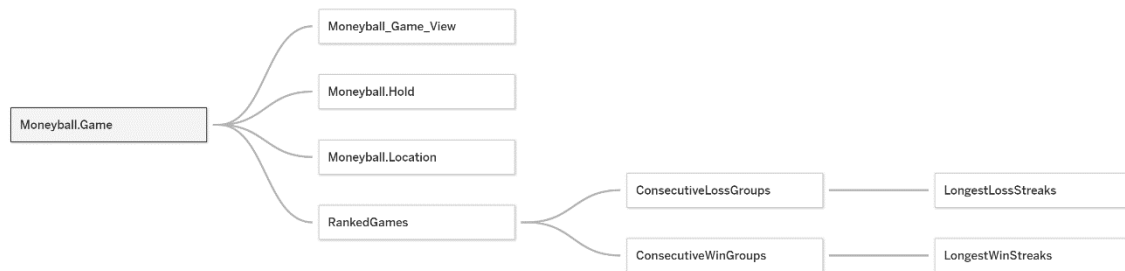
SUM(Opponen..)

Visualization:



Winning Streaks (Q5)

Data Source (for Q5, Q6, Q7):



These connections include all fields required for the last three business transactions.

Configuration:

Columns	Year
Rows	SUM(Longest Win S..

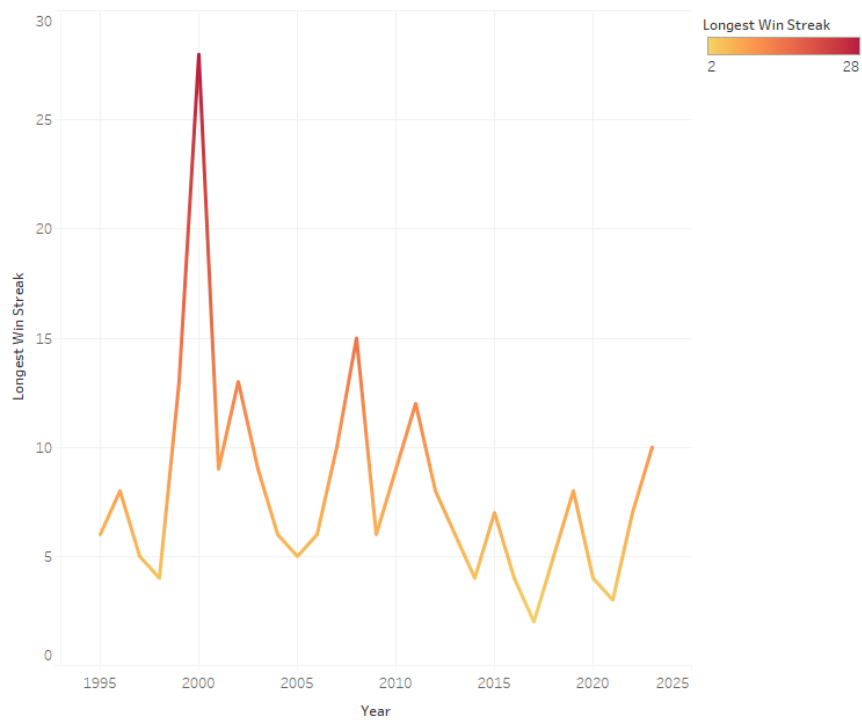
Marks

~ Automatic

Color	Size	Label
Detail	Tooltip	Path

SUM(Longest Win Streak)

Visualization:



Losing Streaks (Q6)

Configuration:

iii Columns	Game Year
Rows	SUM(Longest Loss ..

Marks

~ Line

Color

Size

Label

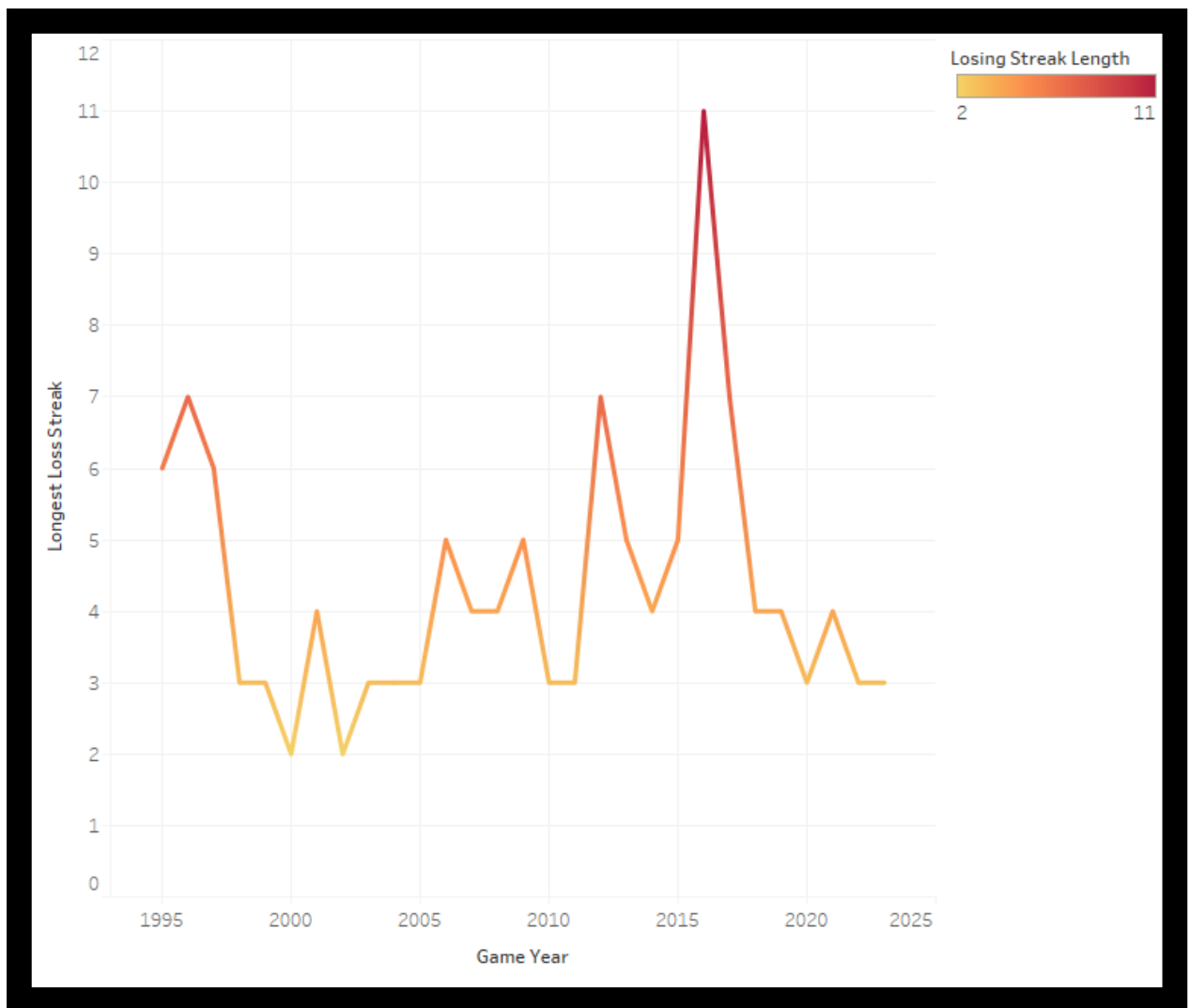
Detail

Tooltip

Path

SUM(Longest Loss Streak)

Visualization:



Performance by time of day (Q7)

Configuration:

Columns	Time Of Day
Rows	AGG(Win Rate)

Filters

HOUR(Gme Time)

Marks

Circle

Color

Size

Label

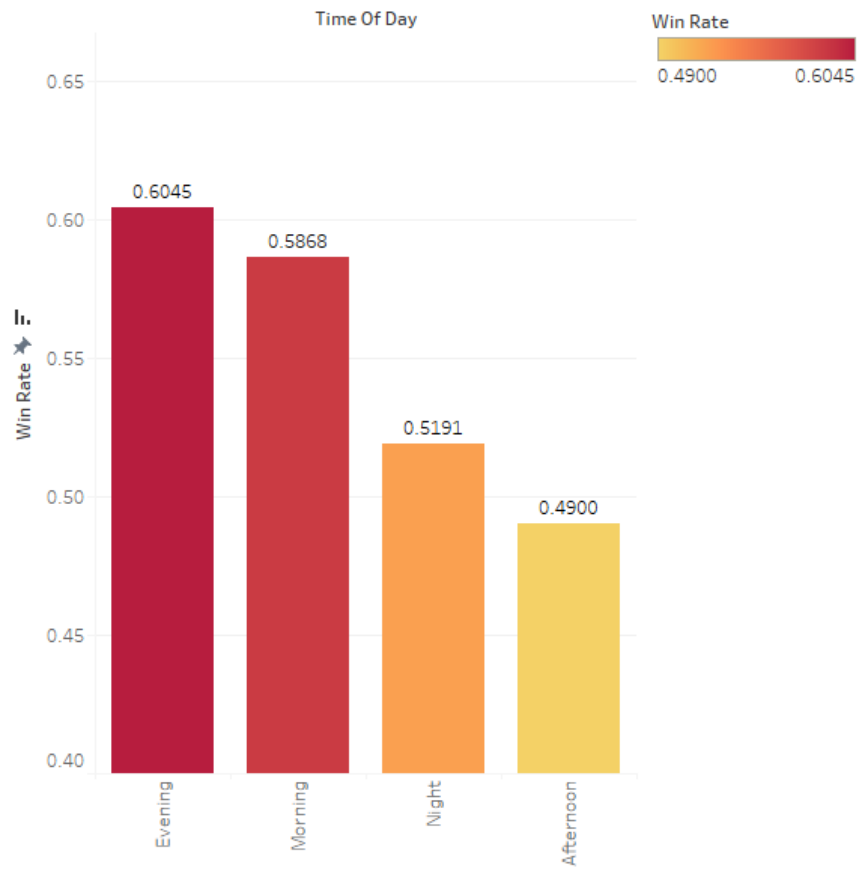
Detail

Tooltip

AGG(Win Rate)

AGG(Win Rate)

Visualization:



Powerpoint:

All the resultant SQL code, outputs and Tableau visualizations attached in the presentation.

