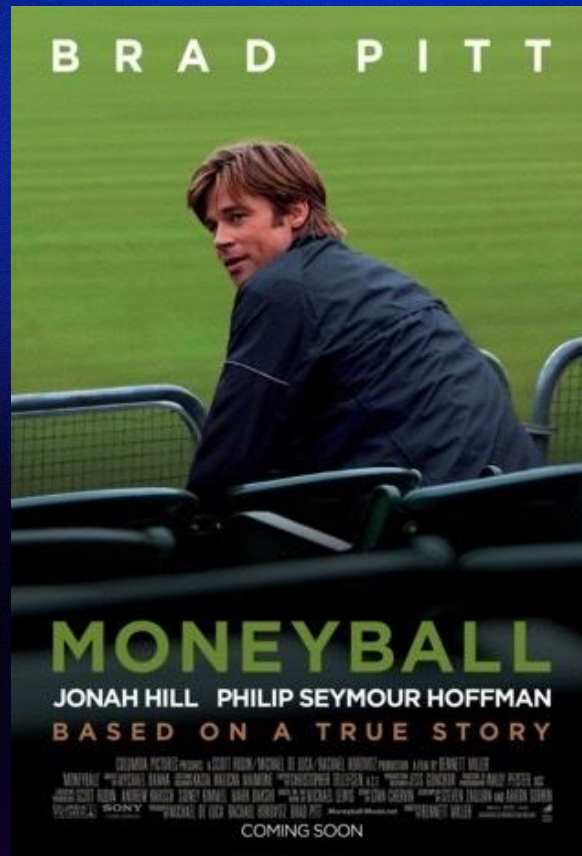


Moneyball: ROI Analysis in MLB



Liam Chaitin

GitHub: github.com/Lpchaitin/moneyball

Elevator Pitch

Winning, but at what cost?

- Who: Scouts, front offices, managers
- Problem: Maximizing ROI with payroll
- Reveal performance-cost imbalances

Why I Fit This Role

Built predictive ROI performance models

Automated SQL and Python pipelines

Delivered insights via Looker dashboard

Focused on business value creation

Strong communicator with analytics skills

API Data Source

- Kaggle: MLB Team Payrolls (2011–2024)
 - Data: Team, Year, Payroll, etc.
 - Relevance: SQL + Python + cloud pipelines
 - Kaggle → .ipynb → PostgreSQL → dbt → Looker

SQL Descriptive Analytics Query

Question: Average Payroll by Year?

Insights: Steep Drop in 2020

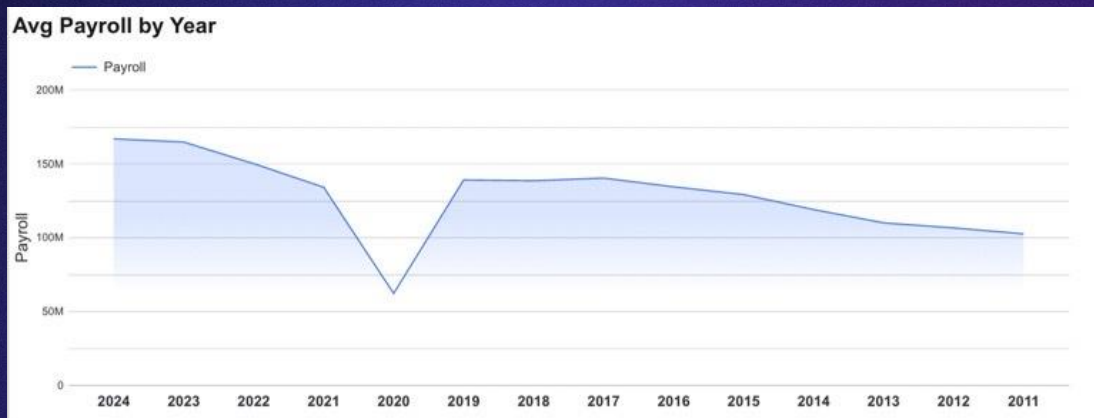
Recommendation: Monitor Spending Growth Rates

Prediction: Avg payroll will level

```
sql_query = '''
WITH payrolls AS (
    SELECT
        "Team Full" AS team,
        "Year" AS year,
        REPLACE(REPLACE("Total Payroll Allocations", '$', ''), ', ', ', ')::numeric AS
        payroll
    FROM sql_project.team payroll records
)

SELECT
    year,
    AVG(payroll) AS avg_payroll,
    MAX(payroll) AS max_payroll,
    MIN(payroll) AS min_payroll
FROM payrolls
GROUP BY year
ORDER BY year;
'''

df = pd.read_sql(sql_query, engine)
df
```



SQL Diagnostic Analytics Query

Question: Largest Payroll Change? Avg Team Payroll?

Insights: Dodgers led both statistics

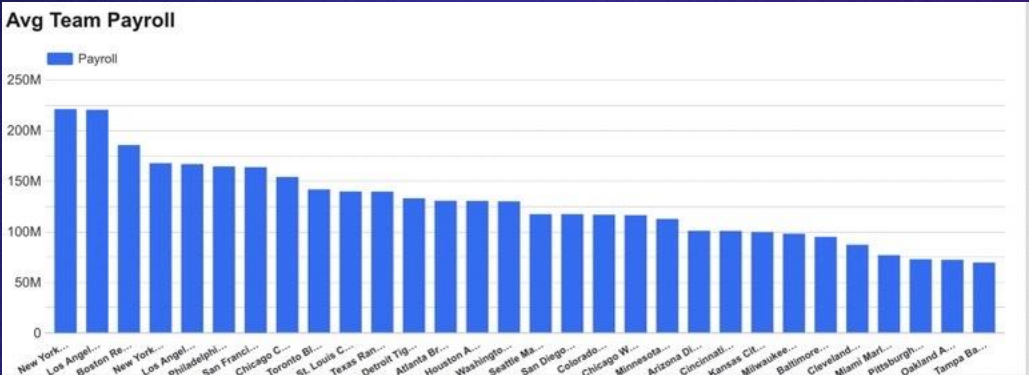
Recommendation: Review correlation with wins

Prediction: Higher payroll != More Wins

```
sql query = '''
WITH payrolls AS (
    SELECT
        "Team Full" AS team,
        "Year" AS year,
        REPLACE(REPLACE("Total Payroll Allocations", '$', ''), ', ', '')::numeric AS payroll
    FROM sql project.team payroll records
),
payroll growth AS (
    SELECT
        team,
        year,
        payroll,
        payroll - LAG(payroll) OVER (PARTITION BY team ORDER BY year) AS payroll change
    FROM payrolls
)

SELECT *
FROM payroll growth
WHERE payroll change IS NOT NULL
ORDER BY payroll change DESC
LIMIT 10;
'''

df2 = pd.read_sql(sql query, engine)
df2
```



Web-Scraped Data Source

- <https://www.baseball-reference.com/>
 - Data: Team, Year, Wins, Losses...
 - Relevance: SQL + Python + cloud pipelines
 - Baseball Reference → .ipynb → PostgreSQL → dbt → Looker

SQL Descriptive Analytics Query

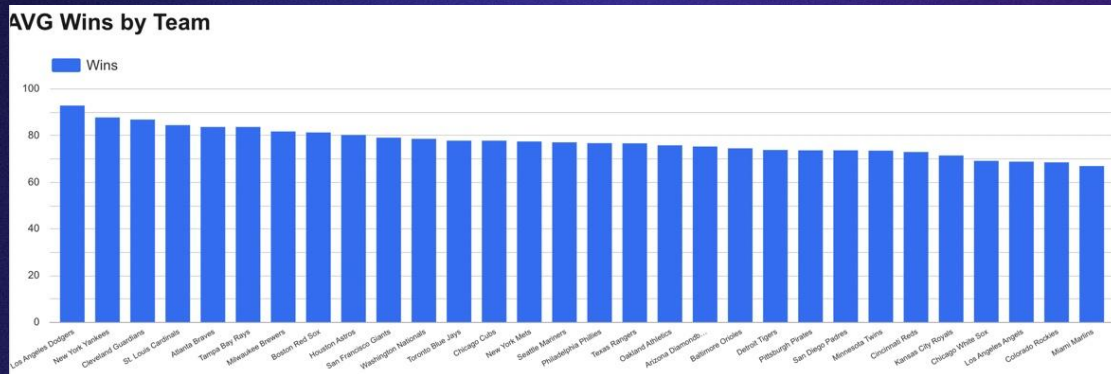
Question: Avg wins? W/L %?

Insights: #1 Dodgers, #30 Marlins

Recommendation: Evaluate Top Performers Strategy

```
sql_query = '''  
SELECT  
    "Team Full" AS team,  
    "Year" AS year,  
    AVG("Win_Percentage") AS avg_win_pct  
FROM sql_project.team_payroll_records  
GROUP BY team, year  
ORDER BY year, avg_win_pct DESC;  
'''
```

```
df = pd.read_sql(sql_query, engine)  
df
```



SQL Diagnostic Analytics Query

Question: Max/Min Losses? Win% Change?

Insights: 2022 Baltimore, 2024 White Sox

Recommendation: Evaluate Baltimore's 2022 Spending

Prediction: Strategic Spending = Competitivity

```
sql query = "WITH win_data AS
(
    SELECT
        'Team Full' AS team,
        'Year' AS year,
        'Win_Percentage'
    FROM sql_analytics_team_data.win_data
    WHERE 'Win_Percentage' IS NOT NULL
)
-- changes AS
(
    SELECT
        team,
        year,
        'Win_Percentage',
        'Win_Percentage' - LAG('Win_Percentage') OVER (
            PARTITION BY team
            ORDER BY year
        ) AS win_pct_change
    FROM win_data
)
-- filtered changes AS
(
    SELECT *
    FROM win_changes
    WHERE
        win_pct_change IS NOT NULL
        AND win_pct_change < 0
)
SELECT
    ROW filtered changes
ORDER BY win_pct_change DESC
LIMIT 10
--
-- pd.read_sql(sql query, engine)
```

	team	year	Win_Percentage	win_pct_change
0	Baltimore Orioles	2022	0.512	0.191
1	Kansas City Royals	2024	0.531	0.185
2	San Diego Padres	2020	0.617	0.185
3	San Francisco Giants	2021	0.660	0.177
4	Boston Red Sox	2013	0.599	0.173
5	Boston Red Sox	2021	0.568	0.168
6	Miami Marlins	2020	0.517	0.165
7	Minnesota Twins	2017	0.525	0.161
8	New York Mets	2022	0.623	0.148
9	Chicago Cubs	2015	0.599	0.148

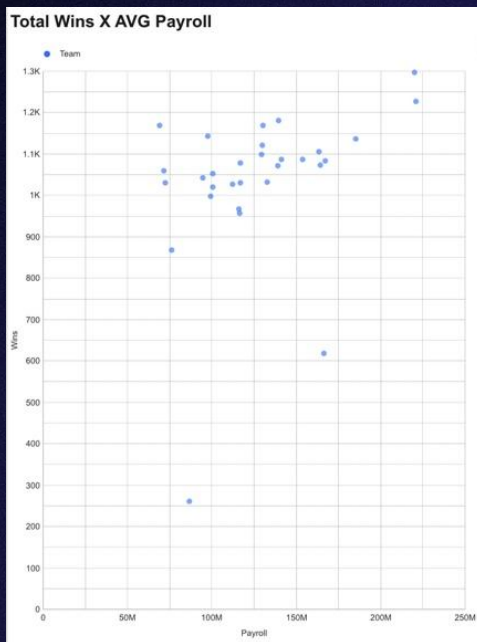
Most Wins & Most Losses in a Season by Team			
	Team	Most Wins & Most Losses in a Season by Team	Losses
1.	Chicago White Sox	107	121
2.	Baltimore Orioles	101	115
3.	Detroit Tigers	95	114
4.	Oakland Athletics	97	112
5.	Houston Astros	107	111
6.	Arizona Diamondbacks	94	110
7.	Washington Nationals	98	107
8.	Kansas City Royals	95	106
9.	Miami Marlins	84	105
10.	Colorado Rockies	91	103
11.	Minnesota Twins	101	103
12.	Texas Rangers	96	102
13.	Chicago Cubs	103	101
14.	Pittsburgh Pirates	98	101
15.	Cincinnati Reds	97	100

ROI ANALYSIS

Insights: Higher Payroll → Higher Payroll/Win

Recommendation: Evaluate Low Payroll, High Win Teams

Prediction: Higher ROI, More Wins



Summary of Statistics

	Win_Percentage	av...	Win_Pe...	Wins	Losses	
1.	New Yo...	3,093,181	220.93...	7.92	1,226	940
2.	Los An...	3,063,691	220.26...	8.46	1,297	869
3.	Boston ...	2,595,770	185.41...	7.26	1,136	1,030
4.	New Yo...	2,345,097	167.50...	6.95	1,082	1,084
5.	Los An...	1,489,761	166.64...	4.09	618	738
6.	Philadel...	2,299,655	164.26...	6.91	1,072	1,094
7.	San Fra...	2,289,710	163.55...	7.13	1,105	1,061
8.	Chicag...	2,154,351	153.88...	7.07	1,087	1,079
9.	Toronto...	1,981,385	141.52...	7.05	1,087	1,079
10.	St. Loui...	1,952,900	139.49...	7.62	1,180	984
11.	Texas ...	1,951,522	139.39...	6.84	1,071	1,096
12.	Detroit ...	1,859,057	132.78...	6.83	1,031	1,130
13.	Atlanta ...	1,824,534	130.32...	7.59	1,169	995
14.	Housto...	1,822,978	130.21...	7.23	1,121	1,044
15.	Washin...	1,816,758	129.76...	7.05	1,098	1,067
16.	Seattle ...	1,639,705	117.121...	6.93	1,077	1,089
17.	San Die...	1,639,680	117.120...	6.74	1,029	1,137
18.	Colorad...	1,631,852	116.560...	6.18	957	1,209
19.	Chicag...	1,625,165	116.083...	6.33	966	1,199
20.	Minnes...	1,572,806	112.343...	6.71	1,027	1,139
21.	Arizona...	1,409,451	100.67...	6.76	1,052	1,114
22.	Cincinn...	1,406,950	100.49...	6.62	1,019	1,147
23.	Kansas...	1,391,402	99.385...	6.43	998	1,188
24.	Milwau...	1,369,049	97.789...	7.35	1,142	1,025
25.	Baltimo...	1,325,929	94.709...	6.89	1,041	1,125
26.	Clevela...	260,613,701	86.871...	1.61	260	225
27.	Miami ...	994,743,261	76.518...	5.89	868	1,134
28.	Pittsbur...	1,014,643	72.474...	6.56	1,029	1,135
29.	Oaklan...	1,007,108	71.936...	6.92	1,059	1,107
30.	Tampa ...	969,621,102	69.258...	7.63	1,169	998

Why Hire Me?

- Modeled ROI using SQL + Python
- Automated data ingestion workflows
- Built insights dashboard in Looker
- Used public sports data sources
- Delivered business-ready recommendations

Thank you!

Contact

Date