



Houston Astros Data Challenge

Effects of Travel on Team Performance

Project Description:

This challenge is to find the effect, if any, that traveling has on team performance within Major League Baseball (MLB). The MLB season consists of 162 games played for each team broken into, most commonly, a series of 3 or 4 games played between the same teams. Often a particular team plays a few consecutive series “on the road” (meaning away from their home stadium and city) in a row which requires traveling between cities and states, sometimes with a day off between games and sometimes not.

For example, based on the schedule shown in Fig. 1, at the start of the 2024 season the Houston Astros will play 4 games in Houston against the New York Yankees, then 3 games in Houston against the Toronto Blue Jays, then have a day off, then play 4 games in Arlington Texas against the Texas Rangers then, without a day off, play 3 games in Kansas City Missouri against the Kansas City Royals, finally returning to Houston Texas to play 3 more games at home. In all the Astros play 20 games in 21 days across 3 different cities. This travel is indicative of a normal schedule for an MLB team.

The image shows a calendar-style schedule for the Houston Astros from March 28 to April 30. The days of the week are listed at the top: SUN, MON, TUE, WED, THU, FRI, SAT. The dates are listed on the left: 28, 29, 30, 31, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20. The schedule is as follows:

DATE	DAY	OPPONENT	TIME	LOCATION
28	THU	NY Yankees	3:10	Houston
29	FRI	NY Yankees	7:10	Houston
30	SAT	NY Yankees	6:10	Houston
31	SUN	NY Yankees	1:10	Houston
1	MON	Toronto Blue Jays	7:10	Houston
2	TUE	Toronto Blue Jays	7:10	Houston
3	WED	Toronto Blue Jays	7:10	Houston
4	THU	Day Off		
5	FRI	Texas Rangers	7:05	Arlington
6	SAT	Texas Rangers	6:05	Arlington
7	SUN	Texas Rangers	1:35	Arlington
8	MON	Texas Rangers	7:05	Arlington
9	TUE	Kansas City Royals	6:40	Kansas City
10	WED	Kansas City Royals	6:40	Kansas City
11	THU	Kansas City Royals	1:10	Kansas City
12	FRI	Texas Rangers	7:10	Arlington
13	SAT	Texas Rangers	6:10	Arlington
14	SUN	Texas Rangers	1:10	Arlington
15	MON	Houston Astros	7:10	Houston
16	TUE	Houston Astros	7:10	Houston
17	WED	Houston Astros	1:10	Houston
18	THU	Day Off		
19	FRI	Washington Nationals	5:45	Washington
20	SAT	Washington Nationals	3:05	Washington

Figure 1: The 2024 Houston Astros Schedule for the start of the season. Here square colored blue or orange represent games to be played in Houston and games with a white background are to be played in the city of the opposing team, denoted by their emblem.

Objective:

The objective here is to determine if the travel schedule of an MLB team is a relevant factor when predicting the performance of that team. This requires

evaluating team performance against some baseline expectation of the true talent level of the team. Teams have different travel requirements (traveling from Houston Texas to Arlington Texas is presumably different than traveling from Houston Texas to Seattle Washington, for example) and schedules have been evolving within Major League Baseball over the years.

The desired outcome is compelling evidence for or against any effect driven by the travel scheduled. Some examples of possible effects could be due to distance travelled for a road trip, consecutive days on the road, whether there is an off day for travel, etc. Judging criteria will be based on the creativity and uniqueness of the approach, the proper and rigorous use of statistical tools and methods to approach the problem, and the level of evidence that is used to justify the final conclusions.

Data Set:

The included data set is based on the open-source data set retrosheet (<https://www.retrosheet.org/>) which offers historical baseball data. The included columns are given in the attached table and includes game level result information for both teams including hitting and pitching stats along with location information the games. The data goes from the 2000 season through the 2023 season.

Prizes:

The winning team will receive unique Astros merchandise.

Additional Resources:

Teams are encouraged to research investigations of this nature within the open baseball research world in addition to resources around estimating team level performance and true skill levels. A good starting resource is the fangraphs library on sabermetric concepts: <https://library.fangraphs.com/principles/expected-wins-and-losses/>.

Column Name	Description	Column Name	Description
game_date	date game was played in yyyymmdd format	home_hr	Number of homeruns hit by the home team.
home_team	abbreviation for the home team	home_fo	Number of field-outs made by the home team.
away_team	abbreviation for the away team	home_so	Number of strikeouts for home team pitchers.
is_day_game	Boolean for the game being played during the day or at night	home_bb	Number of walks by home-team pitchers.
home_score	Final score for the home team	home_hbp	Number of hit-by-pitches for home team pitchers.
away_score	Final score for the away team	away_pa	Number of plate appearances for the away team.
venue	Token for which venue the game was played at	away_1b	Number of singles hit by the away team
venue_name	Name of the venue	away_2b	Number of doubles hit by the away team.
city	City where the game was played	away_3b	Number of triples hit by the away team.
state	US state where the game was played or country is outside of the US.	away_hr	Number of homeruns hit by the away team.
home_pa	Number of plate appearances for the home team.	away_fo	Number of field-outs made by the away team.
home_1b	Number of singles hit by the home team	away_so	Number of strikeouts for home away pitchers.
home_2b	Number of doubles hit by the home team.	away_bb	Number of walks by away team pitchers.
home_3b	Number of triples hit by the home team.	away_hbp	Number of hit-by-pitches for away team pitchers.