

DATA Nutrition Label

About:

This dataset is a set on the NBA fanbases and how they affect how the team plays at home and on the road. The data set is set in the 2022-23 season right after Covid hit so everyone that hasn't seen a game in the past 2 years due to Covid would go to see a game and that would in turn either affect how the team plays on the road or not.

Data Creation Ranges: October 2022 - April 2023

Created By: ESPN and TeamRankings.com

Content: CSV files

Sources: https://www.espn.com/nba/attendance/_year/2023

<https://www.teamrankings.com/nba/stat/win-pct-all-games?date=2023-06-13>

Alert Count 3

Provenance 0

- Misrepresentation 2

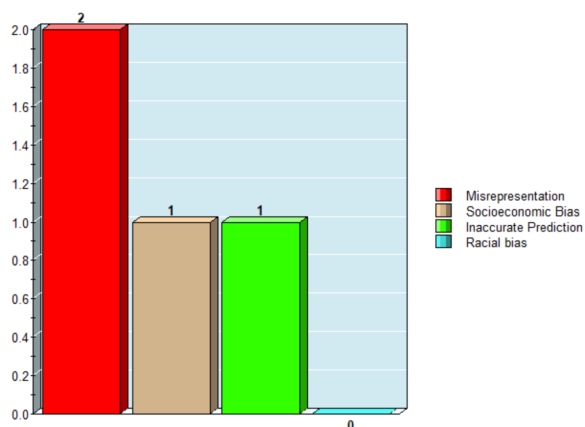
Collection 2

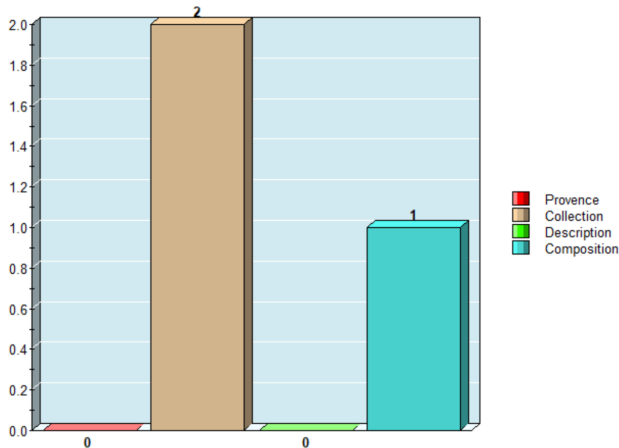
- Socioeconomic Bias 1
- Inaccurate Prediction 1

Description 0

Composition 1

- Racial bias 0





Use Cases:

- 1) Do NBA teams do better with more fans in attendance?
- 2) Has a statistically better team lost due to a lack of attendance?
- 3) Is there a pattern of losses with a certain percentage of the seats not being sold?
- 4) Do the winningest teams get an advantage as they have the most fans in attendance therefore leading to more revenue to spend on the team?
- 5) Does social media help a team's digital presence boost ticket sales which could be beneficial to the team's win percentage?

Badges



Description

1. TELL US ABOUT THIS DATASET.

This dataset is a set of fan base statistics in relationship to the win percentage of each team in the 2022-23 season

2. IS THERE AN INTENDED PURPOSE FOR THE DATASET? WHAT DOMAIN WAS IT DESIGNED FOR?

The purpose of this dataset is the assess whether the implementation of fans affects how the players place

3. **WHAT IS THE LICENSE UNDER WHICH THE DATASET IS MADE AVAILABLE?**
Public Domain
4. **ARE THERE ANY ADDITIONAL RESTRICTIONS UNDER WHICH THE DATASET IS MADE AVAILABLE? ARE THERE ANY LEGISLATION, CODES OF PRACTICE OR GUIDANCE FROM THE JURISDICTION OR DOMAIN IN WHICH THE DATA WAS COLLECTED ABOUT THE USE OF THIS DATA? ARE THERE TASKS FOR WHICH THE DATASET CANNOT BE USED? PLEASE PROVIDE A DESCRIPTION OR GUIDANCE.**
No
5. **ARE THERE TASKS FOR WHICH THE DATASET WOULD BE CAUTIONED AGAINST BEING USED? IF SO, PLEASE PROVIDE A DESCRIPTION.**
No
6. **HAS THE DATASET BEEN USED FOR ANY TASKS ALREADY? IF SO, PLEASE PROVIDE A DESCRIPTION AND LINKS TO PAPERS OR SYSTEMS USING THE DATASET**
No

Composition

1. **DOES THE DATASET HAVE A METADATA REPOSITORY OR DATA DICTIONARY? IF YES, PLEASE PROVIDE THE LINK AND IF NOT, PLEASE EXPLAIN WHAT EACH FIELD MEANS.**
No
2. **IN THE CASE OF EXPLICITLY RELATED INSTANCES, ARE THESE RELATIONSHIPS INDICATED? (E.G., RELATING AN INSTANCE ABOUT A USER TO INSTANCES ABOUT THEIR POSTS)**
No
3. **WHAT DATA DOES EACH INSTANCE CONSIST OF? FOR EXAMPLE, DOES IT CONSIST OF RAW DATA (E.G., UNPROCESSED TEXT OR IMAGES) OR FEATURES? IN EITHER CASE, PLEASE PROVIDE A DESCRIPTION.**
The data is processed and checked by multiple people before released to the public
4. **HOW MANY INSTANCES ARE THERE IN TOTAL (OF EACH TYPE, IF APPROPRIATE)?**
Not sure
5. **IS THERE A LABEL OR TARGET ASSOCIATED WITH EACH INSTANCE?**
No

6. **IS ANY INFORMATION MISSING FROM INDIVIDUAL INSTANCES? IF SO, WHAT IS MISSING AND WHY?**
There is no missing informations
7. **IS YOUR DATASET A SAMPLE? IF SO, WHAT WAS THE SAMPLING STRATEGY USED (E.G. DETERMINISTIC, PROBABILISTIC WITH SPECIFIC SAMPLING PROBABILITIES), AND DOES IT ACCURATELY REPRESENTS THE INTENDED OUTPUT?**
No
8. **WAS ANY PREPROCESSING/CLEANING/LABELING OF THE DATA DONE? IF SO, PLEASE PROVIDE A DESCRIPTION.**
There was as it is an average and not each specific raw data value
9. **IF THE DATA WAS PROCESSED, WAS THE RAW DATA SAVED IN ADDITION TO THE PREPROCESSED/CLEANED/LABELED DATA? IF YES, PLEASE PROVIDE A LINK TO THE RAW DATA.**
No it is only a proceed data. The raw data is not provided in the website
10. **ARE THERE RECOMMENDED DATA SPLITS? IF SO, PLEASE PROVIDE A DESCRIPTION OF THESE SPLITS, EXPLAINING THE RATIONALE BEHIND THEM.**
No
11. **ARE THERE ANY ERRORS, SOURCES OF NOISE, OR REDUNDANCIES IN THE DATASET? IF SO, PLEASE PROVIDE A DESCRIPTION**
There are no errors
12. **DOES THE DATASET INCLUDE INFORMATION FROM UPSTREAM SOURCES? IF SO, NAME THESE SOURCES, THEIR STABILITY, AND ANY KNOWN USAGE LIMITATIONS.**
No
13. **IF THE DATASET INCLUDES INFORMATION FROM UPSTREAM SOURCES, IS THIS DATASET TECHNICALLY DEPENDENT UPON THOSE SOURCES?**
N/A
14. **IS THERE CONFIDENTIAL DATA INCLUDED IN THIS DATASET?**
No
15. **DOES THE DATASET IDENTIFY ANY SUBPOPULATIONS (E.G., BY AGE, GENDER)? IF YES, PLEASE INDICATE THEIR DISTRIBUTIONS WITHIN THE DATASET WHERE POSSIBLE / AVAILABLE.**
No

Provenance

1. **WHO CREATED THE DATASET AND ON BEHALF OF WHICH ENTITY?**

ESPN records the team attendance as it is one of the world's biggest multimedia sports companies and it does this on behalf of the NBA. Team Rankings recorded the win percentage of every team on behalf of the NBA and viewers.

2. HOW IS / WAS THE COLLECTION AND MANAGEMENT OF THIS DATASET FUNDED?

Privately-funded

Collection

1. OVER WHAT TIMEFRAME WAS THE DATA COLLECTED? IF THE DATA CONTINUES TO BE UPDATED, PLEASE INDICATE.

The dataset collects yearly information and is therefore updated yearly. But in this case we are analyzing a data set from 2022-2023

2. WHAT MECHANISMS OR PROCEDURES WERE USED TO COLLECT THE DATA?

This data is collected via stat recorders during NBA games.

3. WHAT IS THE RELATIONSHIP BETWEEN THE DATASET COLLECTOR AND OWNERS OR MANAGERS OF THE DATASET?

ESPN and Team Rankings are sports companies that broadcast the NBA and record stats of it. So essentially they are media partners.

4. IF INDIVIDUALS' DATA IS INCLUDED IN THIS DATASET, DID THOSE INDIVIDUALS CONSENT TO THE COLLECTION AND USE OF THEIR DATA? IF SO, PLEASE DESCRIBE CONSENT PROCEDURE.

N/A

5. IF INDIVIDUALS' DATA IS INCLUDED IN THIS DATASET, WAS THIS DATA ALTERED TO ENSURE HIGHER LEVELS OF PRIVACY? IF YES, PLEASE DESCRIBE ANY PRIVACY PROCEDURES FOLLOWED WITH REGARDS TO THIS DATASET (ANONYMIZATION EFFORTS, PRIVACY PROTOCOLS, SUPPRESSION TECHNIQUES, ETC).

N/A

6. HAS THE DATA BEEN REVIEWED FOR QUALITY?

Yes ESPN checks with the NBA before posting this information to the Public.

7. WERE ANY ETHICAL REVIEW PROCESSES CONDUCTED (E.G., BY AN INSTITUTIONAL REVIEW BOARD)? IF SO, PLEASE PROVIDE A DESCRIPTION OF THESE REVIEW PROCESSES, INCLUDING THE OUTCOMES, AS WELL AS A LINK OR OTHER ACCESS POINT TO ANY

SUPPORTING DOCUMENTATION.

No

Management

1. HOW IS THE DATA BEING MANAGED AT REST AND IN TRANSIT?

Internally

2. HOW CAN THE DATA BE ACCESSED? IF THIS IS GOING TO CHANGE OVER TIME, PLEASE INDICATE THIS.

You can access it through the websites and apps.

3. WILL THE DATASET BE DISTRIBUTED TO ANY OTHER INDIVIDUALS/THIRD PARTIES (E.G. COMPANIES, INSTITUTIONS, ORGANIZATIONS)? IF YES, WHEN WILL THE DATASET BE DISTRIBUTED AND HOW WILL THE DATASET BE DISTRIBUTED (E.G. GITHUB, TARBALL ON THE WEBSITE, ETC)?

These stats are published on many companies' websites just because they are multimedia sports companies.

4. WILL THE DATASET BE UPDATED (E.G., TO CORRECT LABELING ERRORS, ADD NEW INSTANCES, DELETE INSTANCES)? IF SO, PLEASE DESCRIBE HOW OFTEN, BY WHOM, AND HOW UPDATES WILL BE COMMUNICATED TO USERS (E.G., MAILING LIST, GITHUB)?

Data is updated yearly. This is just the 2022-2023 season.

5. IS THERE A MECHANISM THROUGH WHICH AN INDIVIDUAL CAN REQUEST THEIR PERSONAL INFORMATION BE REMOVED?

N/A