Data Analysis & Insights on the NBA

Team 20

Observations in the Dataset

- Observations
  - Many values had inconsistent formatting
  - Many NULL values
  - Team_info_common completely empty
  - Many Unknown Acronyms

Observations in the Dataset

- Solutions
  - Develop a plan on what to clean
  - Create a legend for unknown acronyms
- Used shared document to list our observations

```
#Notes per csv
#game
  #notes about code
  #FGM (goals made), FGA (goals attempted), fg_pct (fgm/fga)
  #fg3m series = 3 pointers
  #ftm = free thrown
  #oerb=# of team offensive rebounds
  #dreb=# of team defensive rebounds
  #reb=# of team rebounds
  #ast=# of team assists
  #stl=# of team steals
  #blk=# of team blocks
  #tov=# of team turnovers
  #pf=# of team personal fouls
  #what is plus_minus_home?
#game_summary
  #many blanks in game sequence?
  #what is game_status?
  #what live pc time?
  #remove broadcast column?
  #what is wh? all column has 1
  #should game code be changed? wasn't listed anywhere else
  #only games after 2000 have network channels
#team_details
  #if string not avalible, should it be left as Null?
#team_history
  #can there be the same nickname for 2 teams?
#team_info_common
  #completely blank
#draft_history
  #some blanks in organization, what does that mean
```

Notes from Observations

Data Cleaning

- Went through data and removed unnecessary tables
- Removed NULL data, and replaced with 0
- Formatted data to proper types
- Created Schema and Tables

```sql
CREATE TABLE team_details (
    team_id int,-- FOREIGN KEY,
    abbreviation char(3),
    nickname varchar(255),
    yearfounded int,
    city varchar(255),
    arena varchar(255),
    arenacapacity int,
    owner varchar(255),
    generalmanager varchar(255),
    headcoach varchar(255),
    dleagueaffiliation varchar(255),
    facebook varchar(255),
    instagram varchar(255),
    twitter varchar(255),
    FOREIGN KEY (team_id) REFERENCES team(id)
);

CREATE TABLE team_history (
    team_id int,-- FOREIGN KEY,
    city varchar(255),
    nickname varchar(255),
    year_founded int,
    year_active_till int,
    FOREIGN KEY (team_id) REFERENCES team(id)
);
```

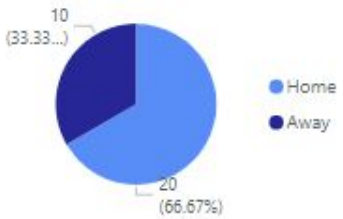
	season	player_id	first_name	last_name	player_name	position	height_wo_shoes	weight	wingspan	standing_reach
0	2001	12033	Adam	Allenspach	Adam Allenspach	C	83.50	259.0	84.50	107.0
1	2001	2240	Gilbert	Arenas	Gilbert Arenas	SG	74.25	199.0	81.50	99.5
2	2001	2220	Brandon	Armstrong	Brandon Armstrong	SG	75.50	188.0	81.50	99.5
3	2001	2203	Shane	Battier	Shane Battier	SF-PF	80.25	229.0	82.50	105.0
4	2001	12034	Cookie	Belcher	Cookie Belcher	SG-PG	75.00	206.0	80.50	99.0
...	...	...	...	...	...	...	...	...	...	...
1197	2023	1641705	Victor	Wembanyama	Victor Wembanyama	C	NaN	NaN	NaN	NaN
1198	2023	1641727	Dariq	Whitehead	Dariq Whitehead	SG	77.75	217.2	82.25	103.5
1199	2023	1641715	Cam	Whitmore	Cam Whitmore	SF	77.75	235.0	80.50	103.5
1200	2023	1630592	Jalen	Wilson	Jalen Wilson	SF	77.50	230.2	80.00	99.5
1201	2023	1631209	Isaiah	Wong	Isaiah Wong	SG	74.50	178.4	78.75	97.5

Create tables script, and cleaning data

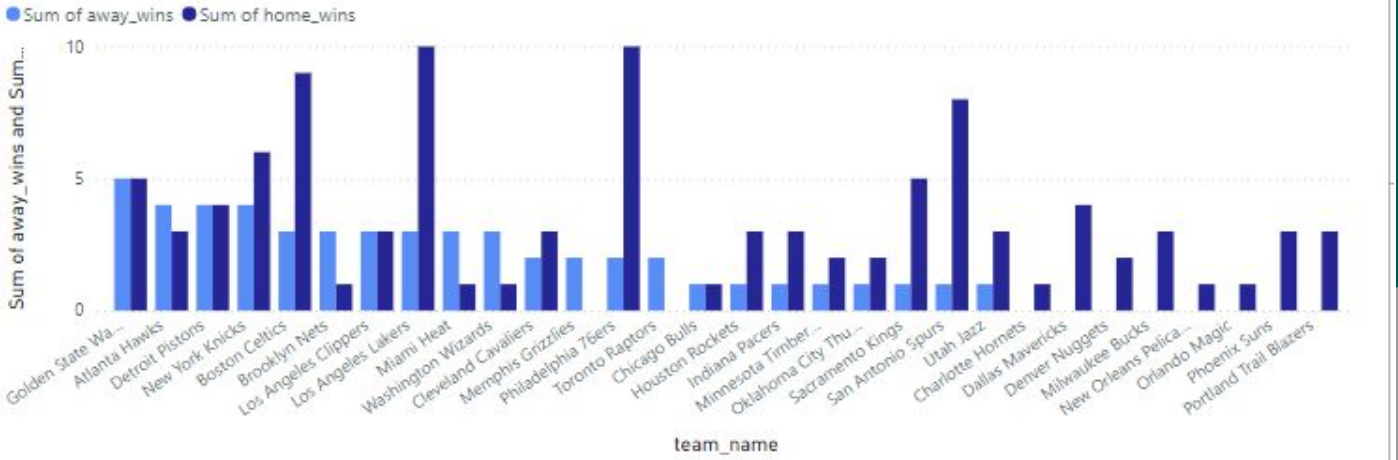
Data Analysis

- Joined tables and created views to find trends
- Graphed trends to see correlation
- Decided to focus on wins and loses compared to home and away games

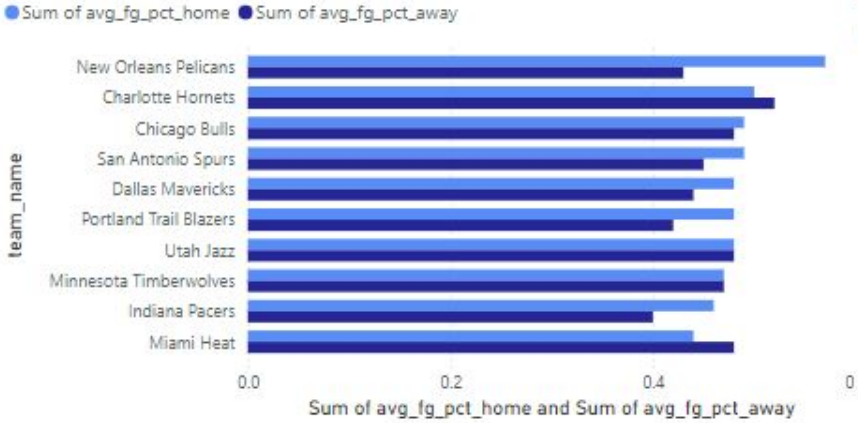
Home Wins VS Away Wins



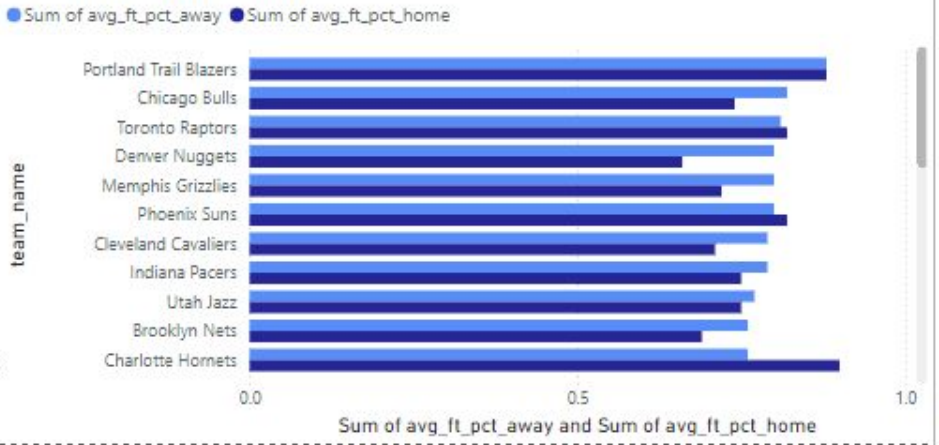
Home Wins VS Away Wins by Team



Avg 2 points Home VS Away by Team



Free Throws Home VS Away



Final Insights

- While observing wins and loses, we saw that majority of the time the teams did better at their home games.
- But when looking at riskier moves and high pressure situations, such as 3 pointers and free throws, teams performed better at away games
- This could be cause of the heightened adrenaline and how some teams perform better under pressure.
- So formulating plays that incorporate these high risk moves, could result in better team scores at away games.