DATA VISUALIZATION PROJECT

MIS 6309.501 - Business Data Warehousing - S17

Instructor:

KEVIN CROOK

Submitted by:

DOMINIC SATISH PETER JEYALES

dxp163130@utdallas.edu

INTRODUCTION TO THE DATASET

DATASET – NBA SHOT LOGS:

We have used a dataset from the open data platform Kaggle. The dataset has around 65535 rows and 21 columns of information on shots taken during the 2014-2015 NBA season, who took the shot, where on the floor was the shot taken from, who was the nearest defender, how far away was the nearest defender, time on the shot clock, and much more. The column titles are generally self-explanatory.

The 2014-2015 NBA season had around 82 regular season matched played among the western conference and the eastern conference teams. The dataset has recorded every single shot that has been attempted by a player in those regular season matches. The dataset does not only contain the dimensions and metrics about the player and the shot, it gives a holistic view of almost everything that is relevant to each shot. Details like defender distance, shot clock time remaining provide a different level of information that can be utilized for interesting analytics.

The dataset suggests that about 144 players have attempted a shot in the 2014-2015 regular season. About 513 matches have been played in the regular season. Of the 65535 shots that have been attempted around 45% of them have been converted. And the season has seen 464 different shooter closest defender combination while a shot was being attempted. And as an average, players have taken two dribbles before hoicking a shot signifying the shoot first mentality the league is trending towards.

ATTRIBUTE NAME	DESCRIPTION
GAME_ID	Unique id for each match
MATCHUP	Date and teams competing
LOCATION	Away or Home
W	Win or Lose
FINAL_MARGIN	Difference in the final score
SHOT_NUMBER	Order of shots made in the game
PERIOD	Shot made in which quarter
GAME_CLOCK	Time remaining in each quarter
SHOT_CLOCK	Time remaining in the 24 second shot clock
DRIBBLES	Number of dribbled made prior to the shot
TOUCH_TIME	Time ball was held before the shot
SHOT_DIST	Distance from the rim when the shot was attempted
PTS_TYPE	Two pointer or a three pointer
SHOT_RESULT	Made or missed
CLOSEST_DEFENDER	Who was the closest defender when shot was attempted
CLOSEST_DEFENDER_PLAYER_ID	Unique id for the defender
CLOSE_DEF_DIST	Closest defender's distance from the shooter
FGM	Field goal made or missed
PTS	Points awarded for the shot
PLAYER_NAME	Player who attempted the shot
PLAYER_ID	unique id for the player

DATA CLEANSING

The dataset had 3159 missing observations for the shot clock field. All the other fields had zero missing observations and were in a format that was ready to be used for the visualization. We have not performed any data cleaning. The missing observations have not been deleted as the other columns provided valuable information. Only in visualizations where the shot clock metric has been utilized did we filter out the blank records.

ANALYTICAL QUESTIONS – HIGH LEVEL

The dataset is feature rich with around 21 columns with around 9 measures and remaining dimensions. The dataset is very non-ambiguous and has names of all the players and competing teams along with their unique ids to enable creating clear quality visualizations.

Unlike a standard shot log dataset which covers only the offensive end of the game shots, the dataset that we have used also has fields like closest defender, closest defender player id and closest defender distance to accurately gauge the kind of defensive pressure the offensive player was under while attempting the shot. By including metrics like shot clock and game clock we are also able to understand the kind of mentality the player would have been in while attempting the shot. Coupled with the field touch time we can tell how much of an impact the defender had on the offensive player's shot attempt.

The dataset not only details out the number of shots that have been attempted it also has fields like points type and distance from the basket to find interesting analytics as to how the game has evolved from a low post strategy employing organization to a happy go lucky jump shooting one. It helps one to understand the number of three pointers and long two pointers that have been attempted in a season. These are generally considered to be low percentage shots. But the league has gotten so comfortable with this approach as it also helps in spacing out the floor for the offense to work on. We have made interesting analysis on the number of three pointers each player has made in the season and have visualized the top ten players and their respective shots made for the season.

The dataset has fields that can divulge information on the kind of mental impact things like playing away and in crunch time situations has on the players. The location fields tell you if the shot was made by the player playing in his home court or away. Fields like period, game clock and final margin can give you insights about when a certain type of shot was attempted. Playing in the second and third quarter and playing when the score is beyond the opponents reach relieves the offensive player from pressure and help him shoot away the night and build his confidence. So, by using the right filters on these fields we will be able to gauge the performance of players in pressure time scenario.

The dataset can help understand the playing style of certain players. The dataset comprises of fields like touch time and dribbles which hold information about the total seconds that a player had possession of the basketball prior to shooting it. The dribbles field can tell you if the player dribbled the ball or was holding to it desperately to find a free teammate or to shoot it. We have used these attributes to provide a visualization on players who end up possessing the ball for a prolonged time before shooting them.

The metrics available in the dataset provide a platform to create calculated fields like percentage of field goals made. A player's efficiency can be impacted by the high number of shot that are being attempted. A high shot attempt making player's efficiency should not be compared with the efficiency of a low shot attempt taking player. Fields like field goal made provides information on the number of shots attempted by a player. Filtering the field to our necessity and coupling the information with the calculated field that we can create we will be able to gauge the kind of efficiency a high shot attempting player.

The dataset has a field W which denotes if a game was won or lost by the player attempting the shot. The information is available for all the records. This could provide curious insights as to the kind of impact a player has on a game. We can couple the W field with player name, field goal made and the calculated field percentage of field goal made to provide analysis about how a player fares in winning matched and in losing matches. This could tell the kind of impact a player has on his team.

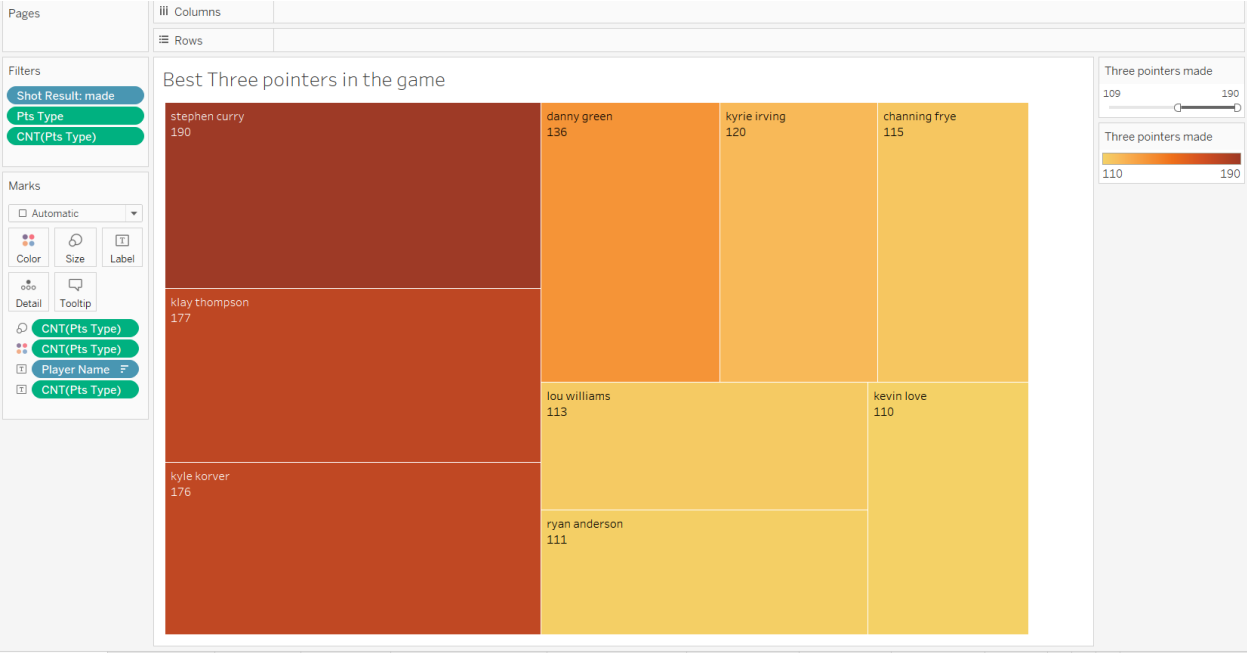
The final margin field can be used along with player name and field goals made field to understand if a player taking certain number of shots is beneficial or detrimental to the team. This can help frame strategies for a certain match up and determine the kind of plays that can be run for a player who the data shows to be beneficial to the team when attempting more shots and curb down on the plays for a player who is detrimental. Usually a team benefits when its star player is at the center of things this analysis could back up that statement with data and visualizations.

FIVE ANALYTICAL QUESTIONS

- Did Stephen Curry deserve the 2014-2015 season NBA MVP award?
- Who is one of the best two way players in the league for the 2014-2015 season?
- Do clutch time players need to have the ball in their hands for a long time to contribute?
- Should the NBA move to a more three pointer oriented style of play?
- Does playing away have an impact on top NBA player's performance?

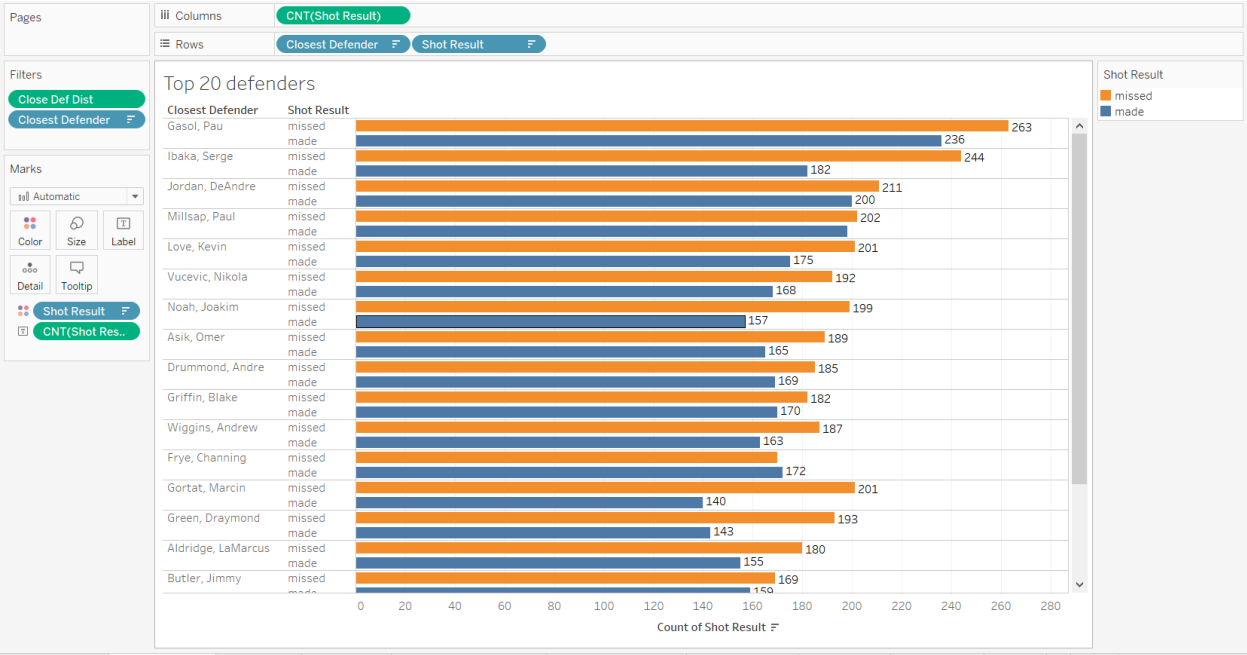
DATA VISUALIZATIONS

BEST THREE POINTERS IN THE GAME:



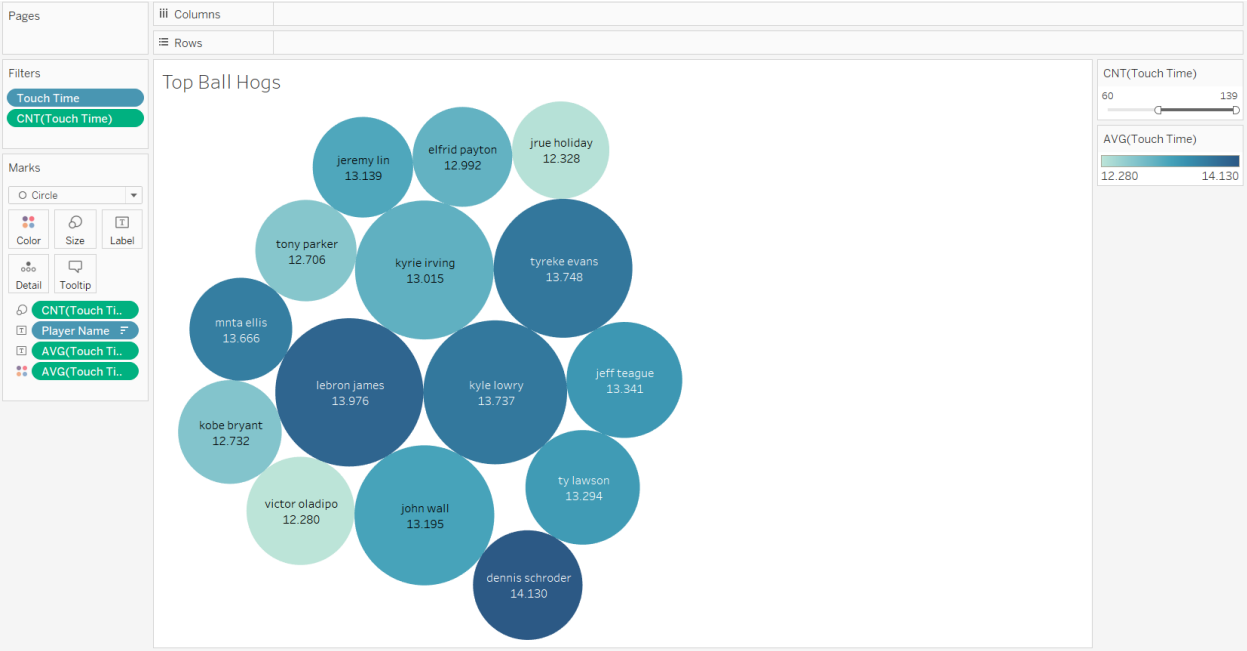
The above visualization of the best three pointers in the game was created by filtering the dataset to display only the records where the shot result is 'made' and the point type was 3. The player name dimension was added to the label mark along with the count of the number of shots made by the player. The number of three pointers made filter has been added to the visualization and made available at the right to enable the user to control the limits of the three pointers made count. We have set the filter to show all the players that have made above 109 three pointers in the 2014-2015 season. The color used in the tree map is orange and it varies from light to dark based on the number of three pointers made ranging from minimum to maximum.

TOP 20 DEFENDERS:



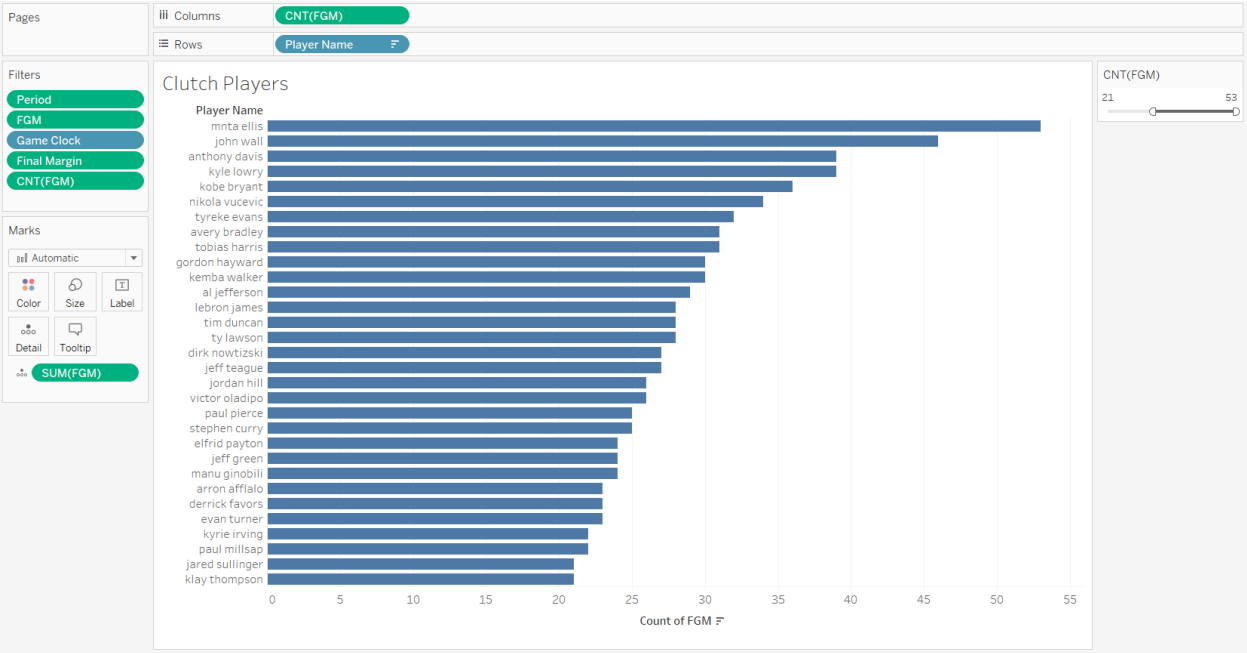
The above visualization of the top twenty defenders was created by filtering and displaying only the data where the closest defender distance was less than 10. Closest defender and shot result dimensions were used as rows and the count of shot results was set as the column to display this stacked bar visualization. The count of shot result has also been added as a label mark to give the user an idea of how many shots the defender has contested and how many shots were made or missed while he was defending. The shot result has been added to the color mark and the orange bar signifies the missed shot count and the blue bar signifies the total number of shots made by a player while being guarded by the defender in the visualization.

TOP BALL HOGS:



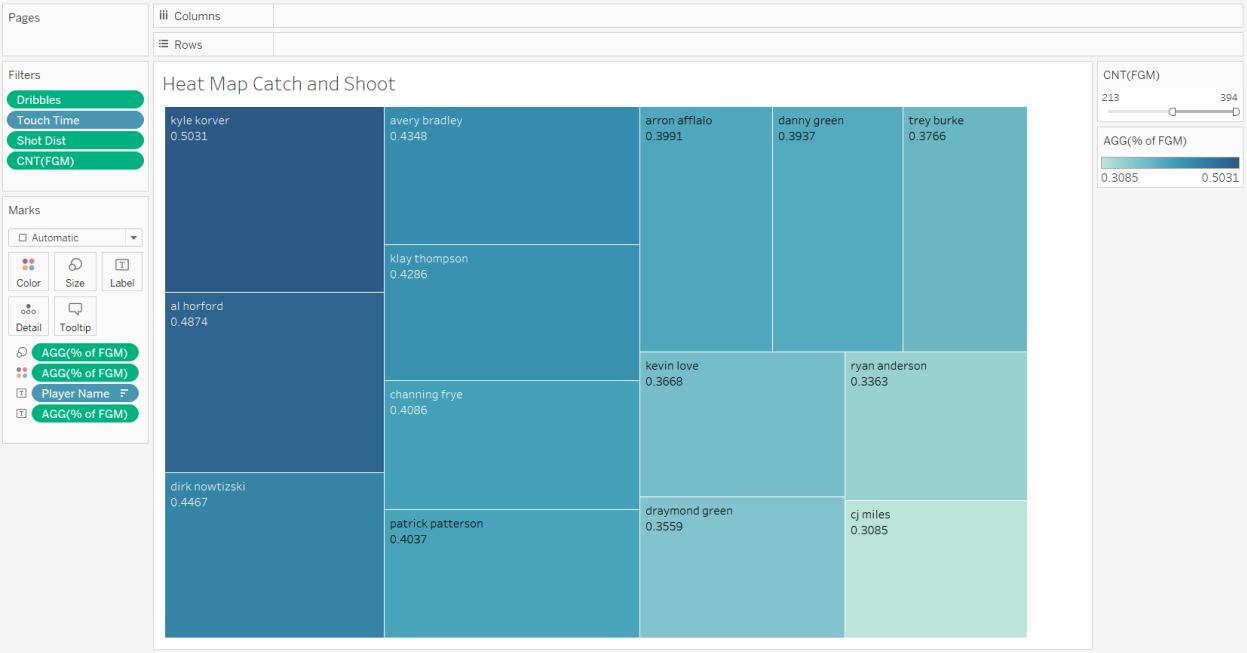
The above visualization of the top ball hogs in the 2014-2015 NBA season was created by filtering the data based on the touch time. All data where a player's touch time was more than 10 seconds were selected and the rest was filtered out. The player name dimension was added to the label mark along with average touch time to provide the user with visual statistics about the player. The average touch time for each player when he has kept the ball to himself for ten or more seconds has been added to the color mark and set to range from light to dark blue based on the usage. The filer count of touch times greater than ten seconds was each player was added to the visualization to give the user control over the number bubbles they would like to see in the visualization.

CLUTCH PLAYERS:



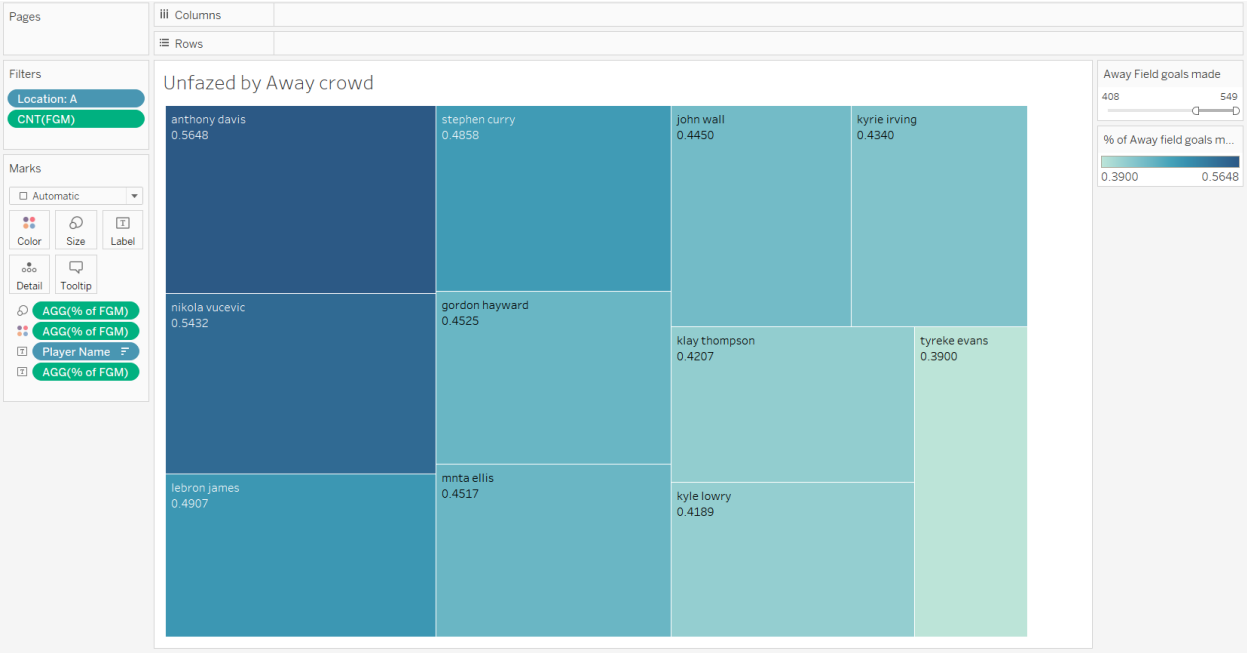
The above visualization of top clutch players in the 2014-2015 NBA season has been created by filtering records based on multiple criteria. Only the shots that were made has been considered. The shots should have been attempted in the fourth quarter when time remaining was less than five minutes or in overtime. The final margin field has also been added to the filter to consider only the shot that were made in a neck to neck match and those in blowouts were disregarded. The player name dimension has been set as the row and number of field goals made by the player has been set as the column. The total number of field goals made by a player has been set as a filter for the user to control and get the range as needed. The top clutch players are displayed in a descending order based on the count of shot they made in clutch time scenarios.

TOP CATCH AND SHOOT PLAYERS:



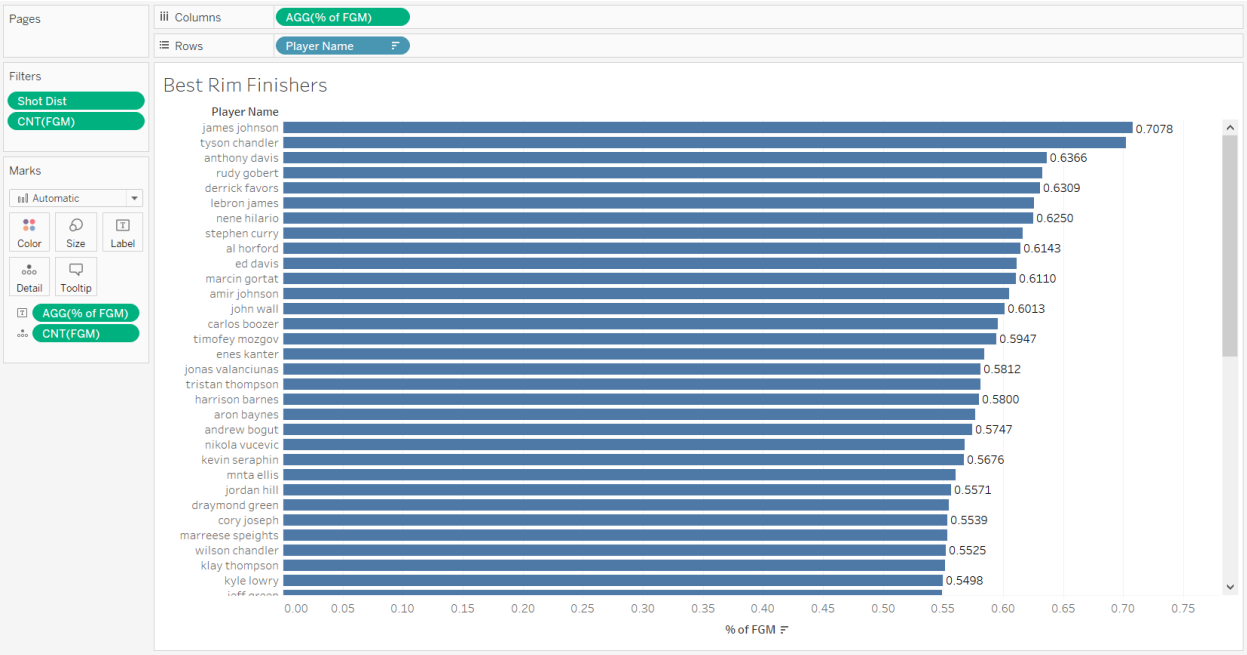
The above tree map visualization of top catch and shoot players has been created by filtering the data to have only the records where the touch time for a shot was between 0 and 1 seconds. A new calculated field % of field goals made has been created for this visualization and added in the size, color and label marks. The records were also filtered to only retain data where the number of dribbles made by the player prior to taking the shot was zero. Also, we have taken into consideration only shots that were made at least fifteen feet away from the rim. The tree maps have the players name and their respective field goal percentage for catch and shoot scenarios. The block size and color are based on the field goal percentage for the individual player. The count of field goals made filter option has been made available to the user to control the visualization. We have set the filter to display the top fourteen players in terms of field goal percentage.

PLAYERS UNFAZED BY AWAY CROWD:



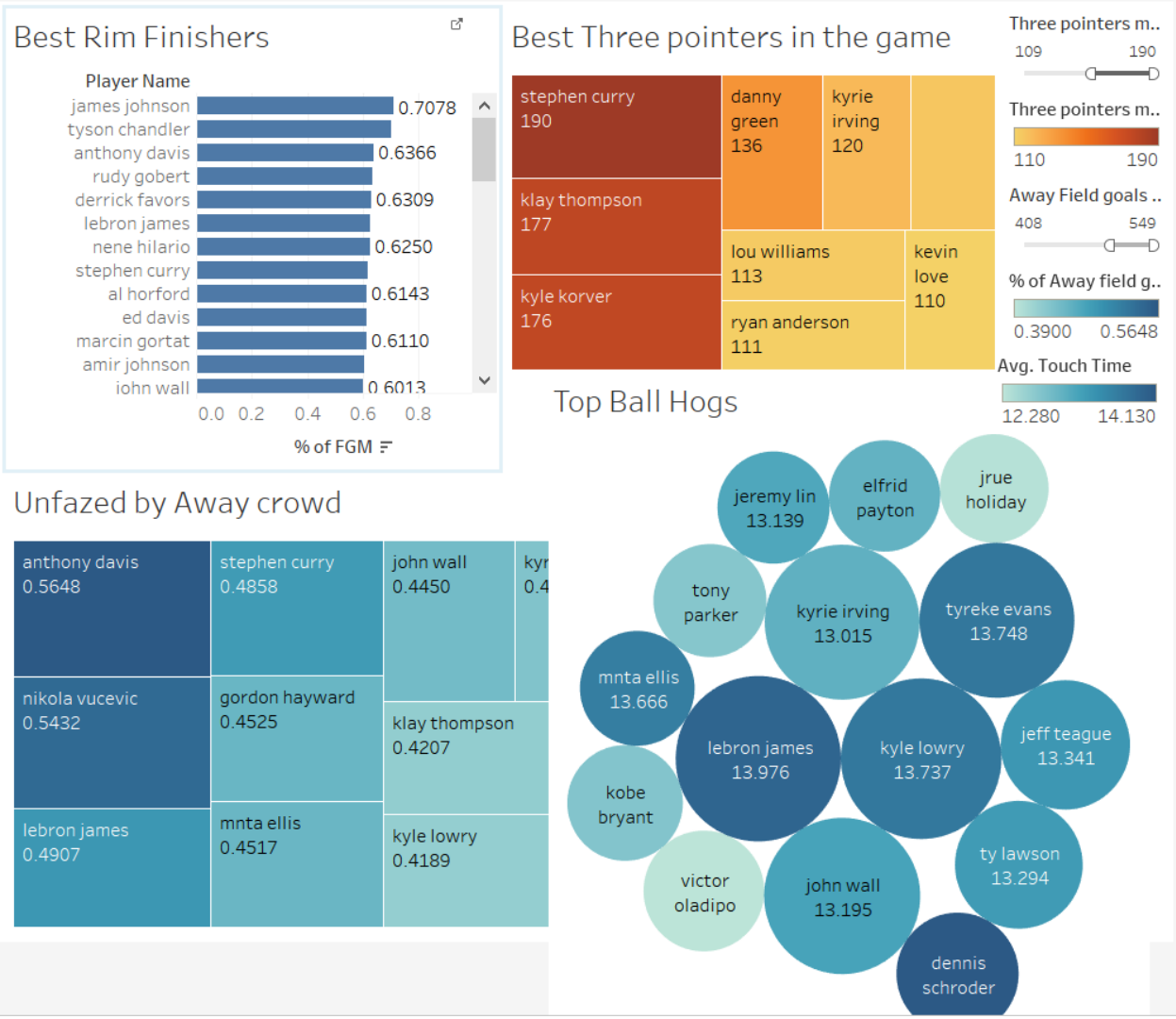
The above tree map visualization of players unfazed by away crowd has been created by filtering the data to consider only the records where a shot has been made by a player playing an away match. Once the filter has been set a threshold has been set to consider only those players who have made a minimum of 408 made shots in the season playing away to disregard players who might made a single shot or a handful of shots and pique their field goal percentage. The aggregate of field goal made has been added in the color, size and label marks. The player name has also been added to the label mark for ease of understanding the visualization. The away field goals made filter has been made available to control by the user to get details only for the information needed by them.

BEST FINISHERS IN THE PAINT:



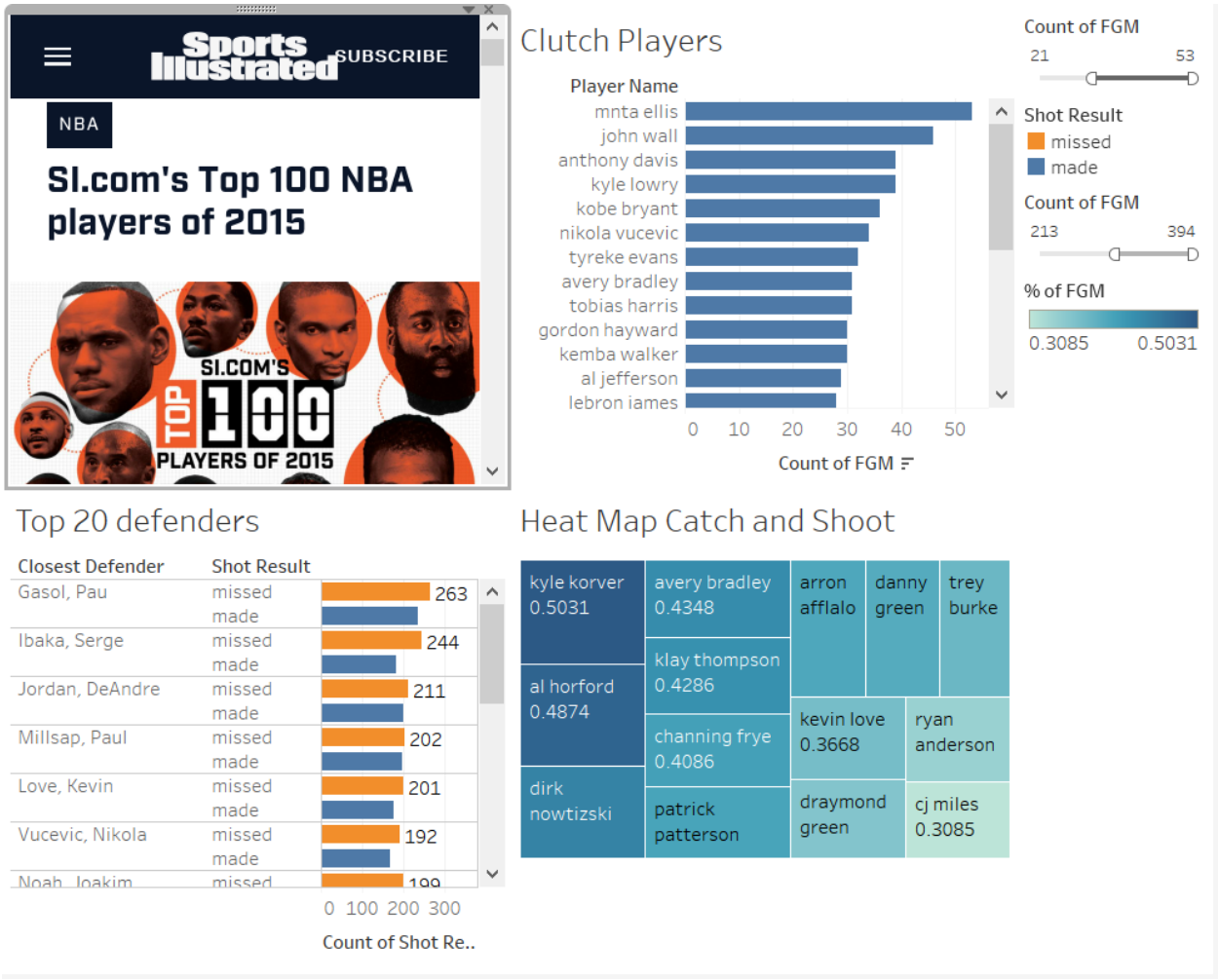
The above visualization of the league's top finishers in the paint for the 2014-2015 season has been created by filtering the data to consider only the shots that have been attempted from at a distance of 10 feet or below from the basket. The count of field goals made by the players in paint has been set to be above 200 to exclude outliers. The player name dimension has been set to be the row for the visualization and the percentage of field goal made by the individual players has been set as the column. The horizontal bars have the players name, count of field goal made and the percentage of field goal made, it can be seen when the user hovers over the bars. The field goal percentage information has been set to the label mark and can be seen to the right of the bars.

DASHBOARD1:



The first dashboard created for the project has four sheets collaged to create a visual representation of a player's impact in the game of basketball for the 2014-2015 NBA season. The sheets have details like best rim finishers, best three pointers in the game, players most unfazed by away crowd and the top ball hogs of the season. Unsurprisingly a few of the names find a place in more than one visualization proving their offensive efficiency.

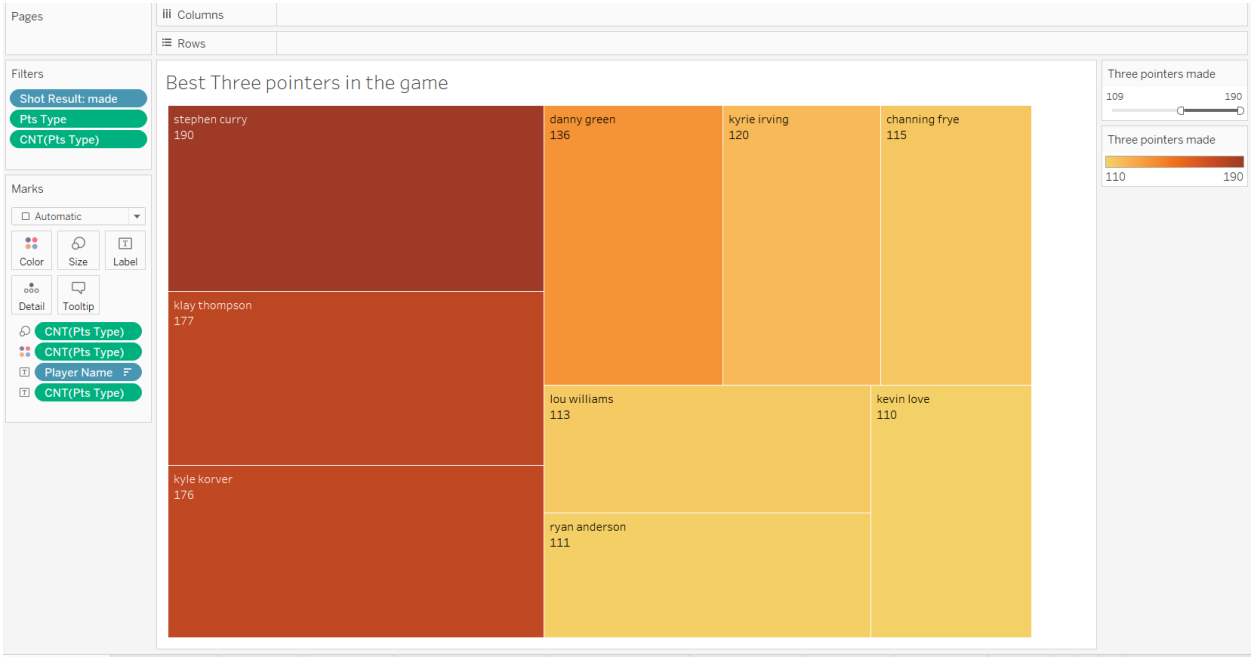
DASHBOARD 2 WITH MASHUP:

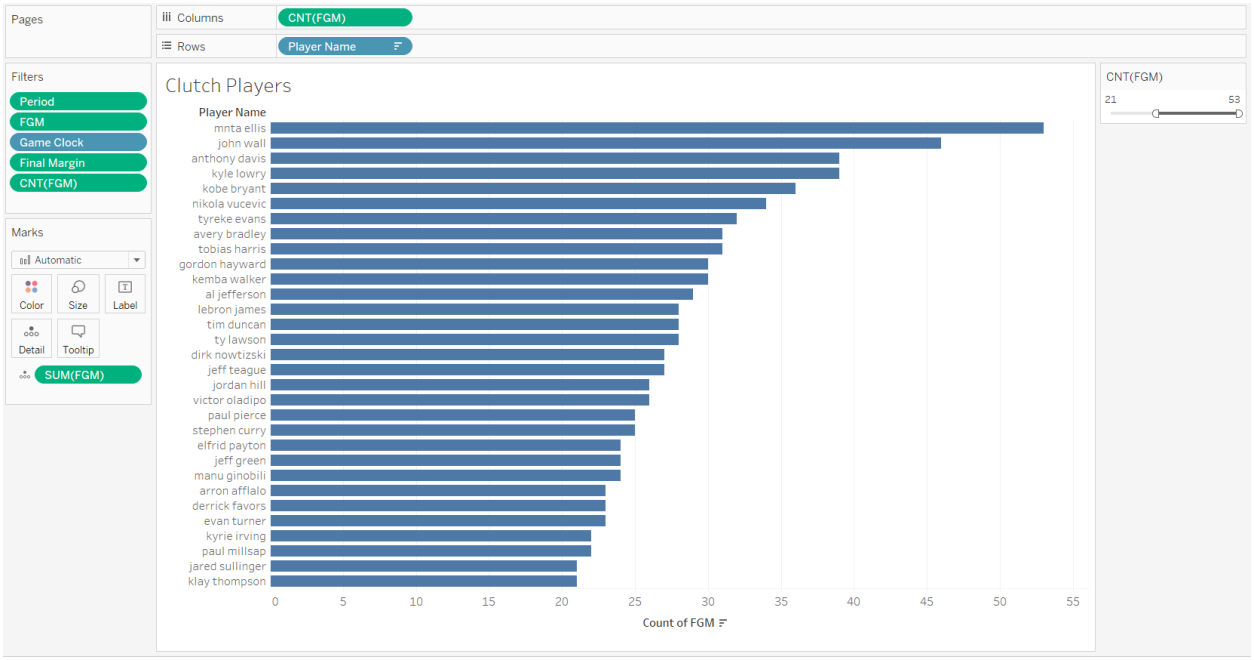


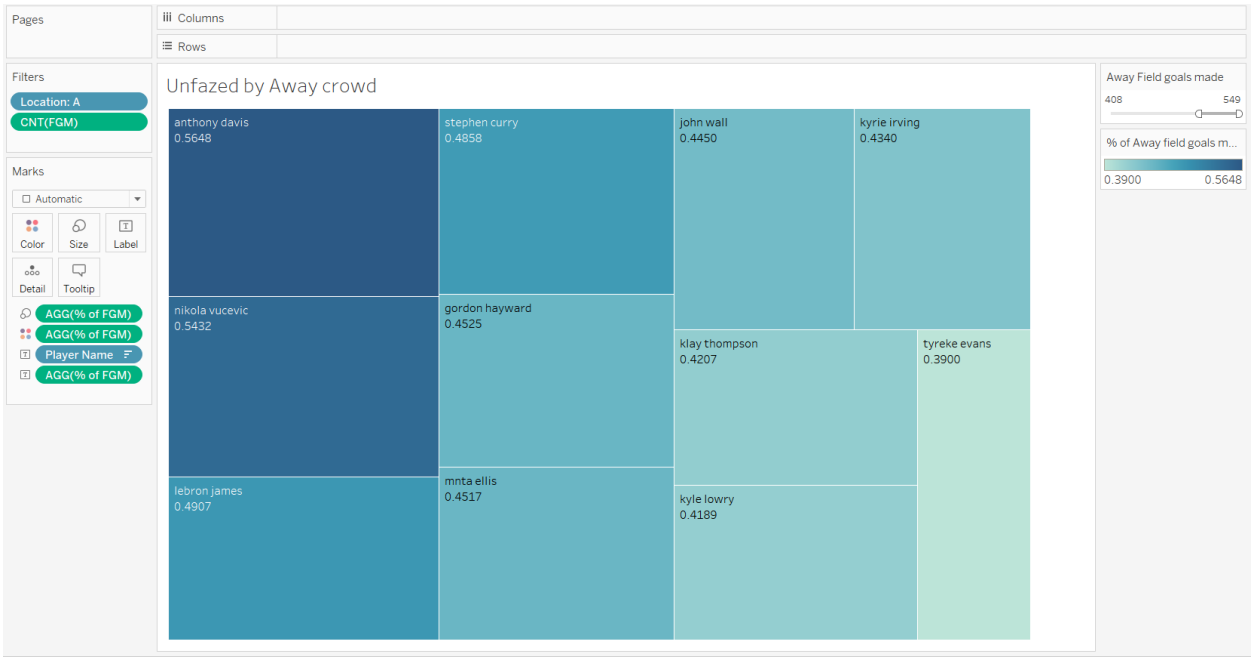
The second dashboard created for the project has three sheets which displays visualizations that hold information about the clutch players, top 20 defenders and best catch and shoot players of the 2014-2015 NBA season. The dashboard also has a mashup in the form of a webpage from sports illustrated site and has a list of the top 100 NBA players of 2015 season. This web page compliments the analytics we have derived from the visualizations created for the project.

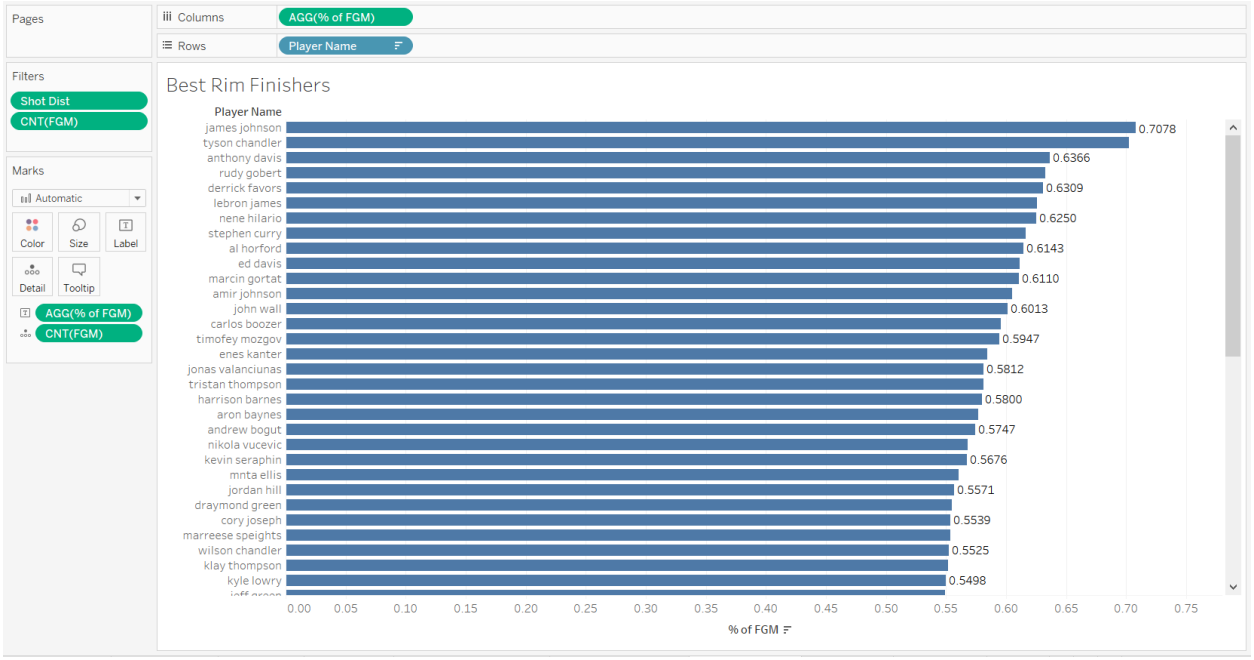
FIVE ANALYTICAL QUESTIONS – EXPLAINED

- Did Stephen Curry deserve the 2014-2015 season NBA MVP award?



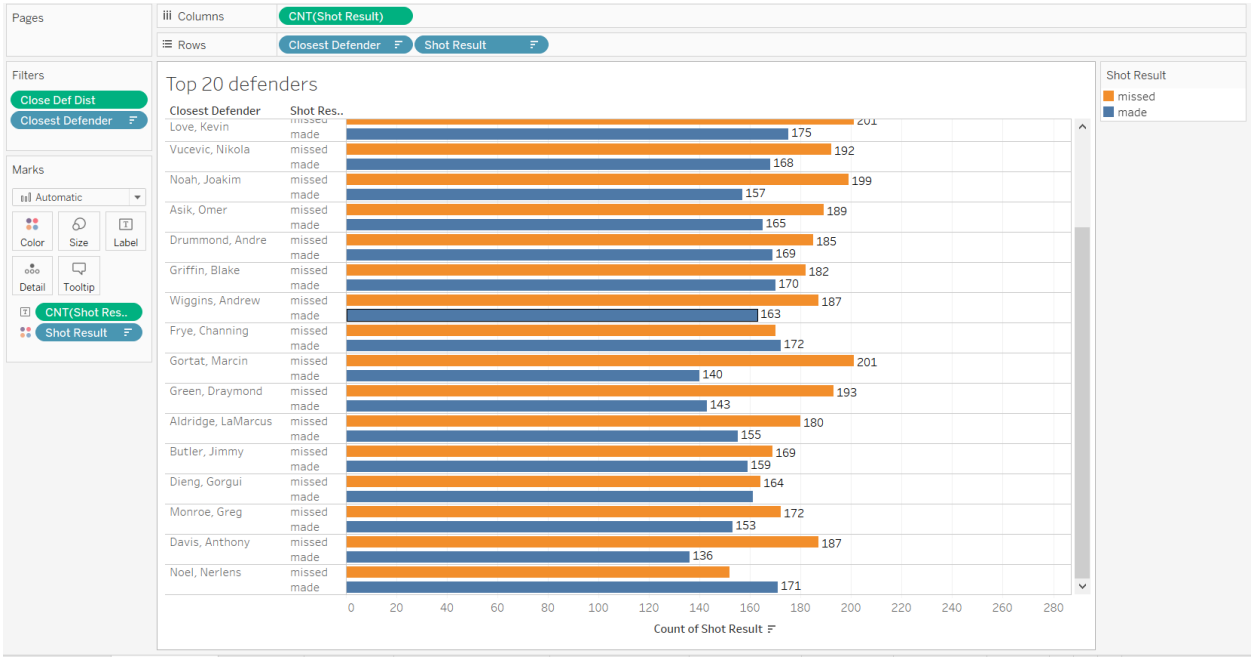


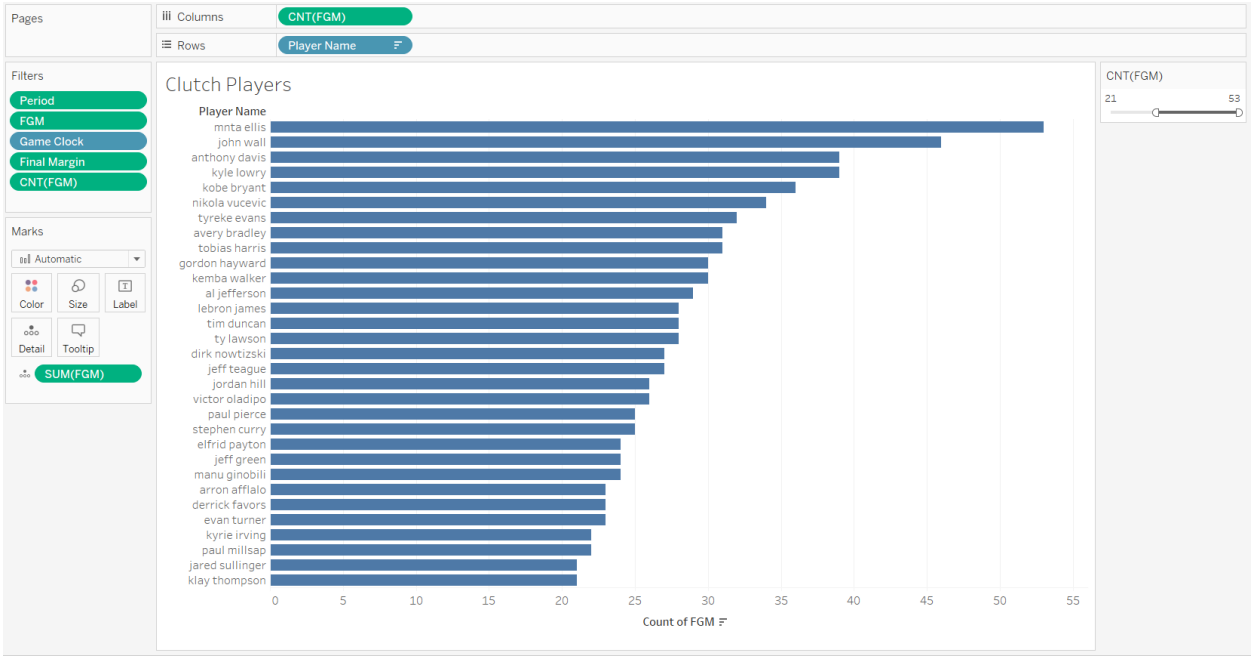


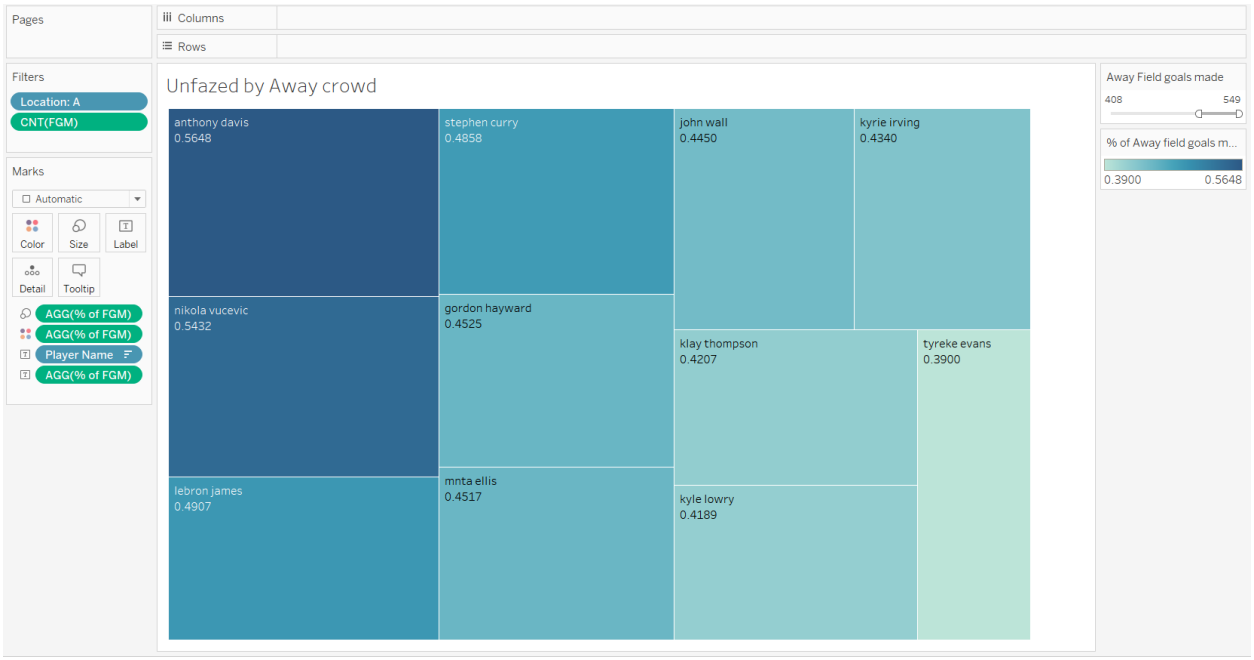


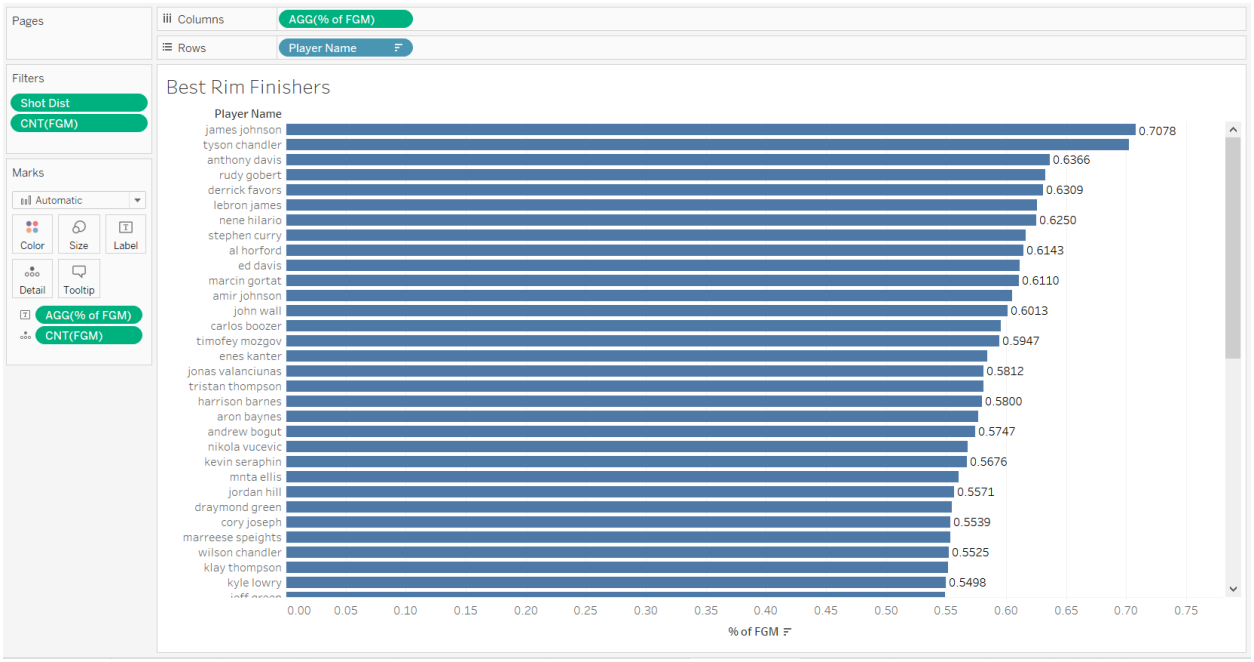
Stephen Curry is the top most player in the three pointer made list, he finishes in the top ten in the players unfazed by away crowd and best finishers in the paint list along with a top twenty place in the clutch players list. For his offensive prowess Stephen Curry is indeed a deserving MVP winner for the 2014-2015 NBA season.

- Who is one of the best two way players in the league for the 2014-2015 season?



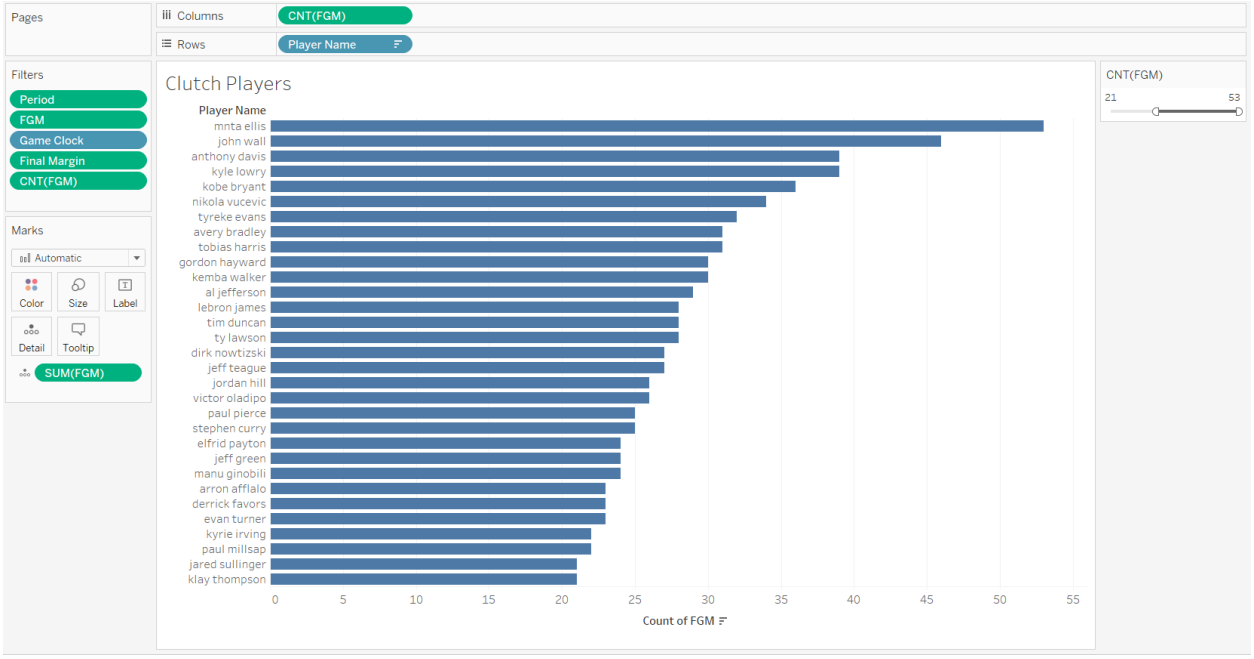


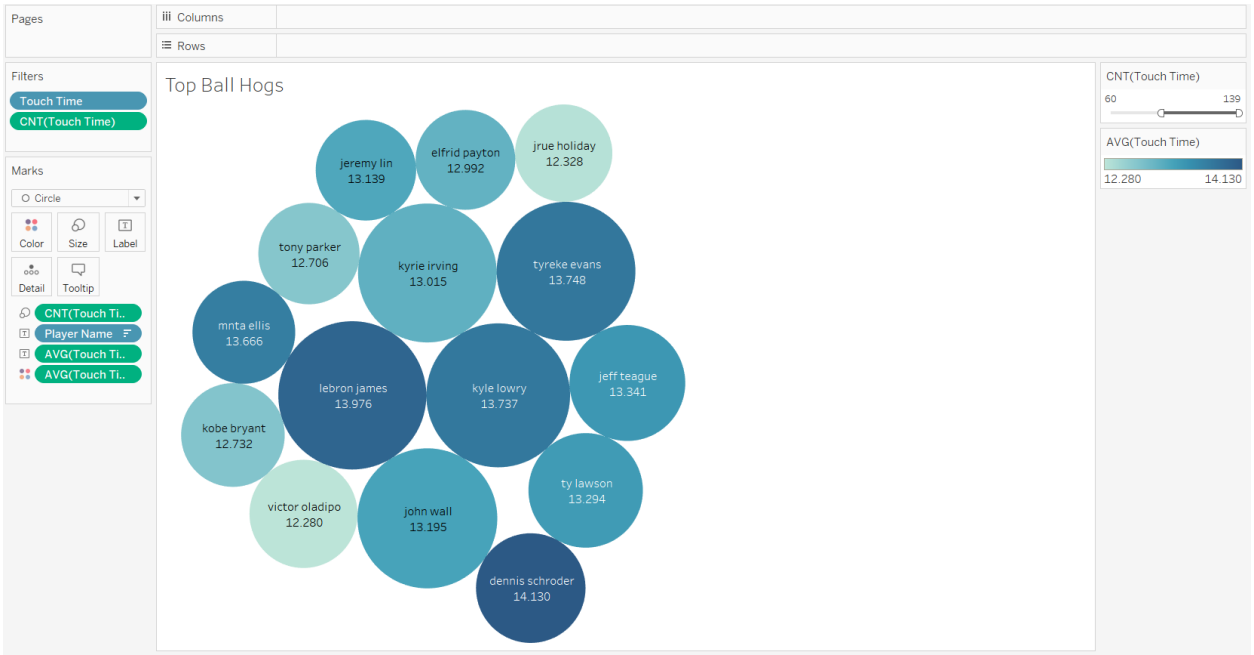




Anthony Davis is the only player in the top twenty defenders to also find a place in the top five for clutch players, unfazed by away crowd and the best rim finishers list thereby validating his claim to be one of the best two way players in the league.

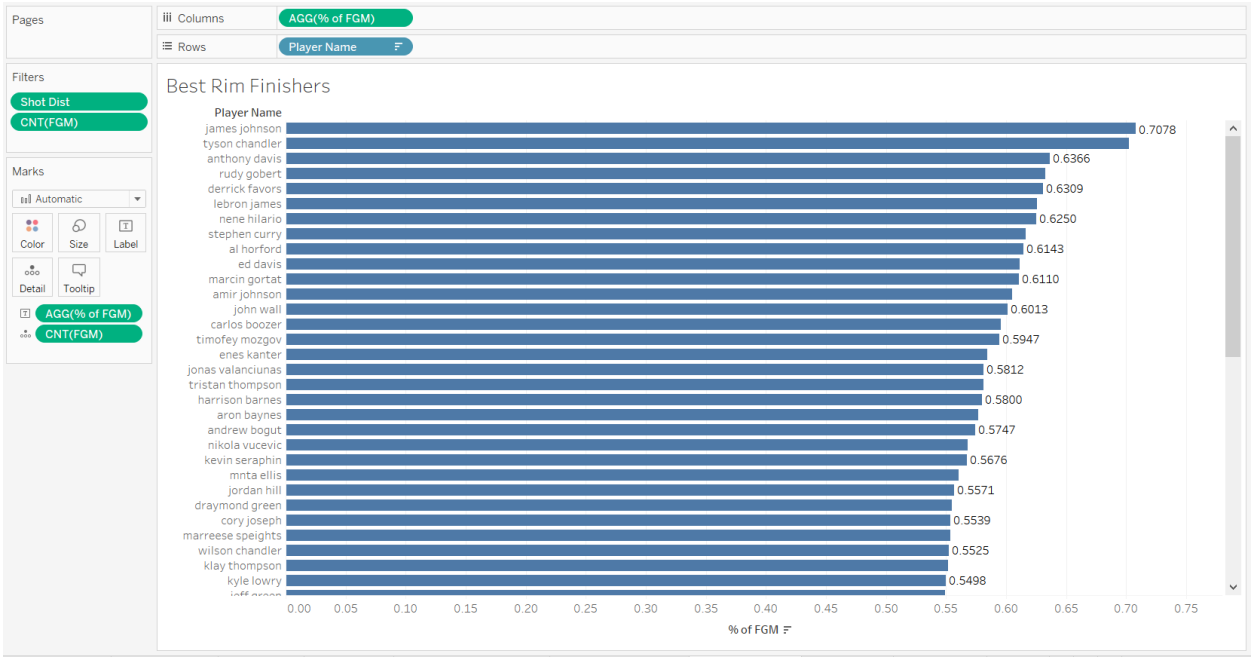
- Do clutch time players need to have the ball in their hands for a long time to contribute?





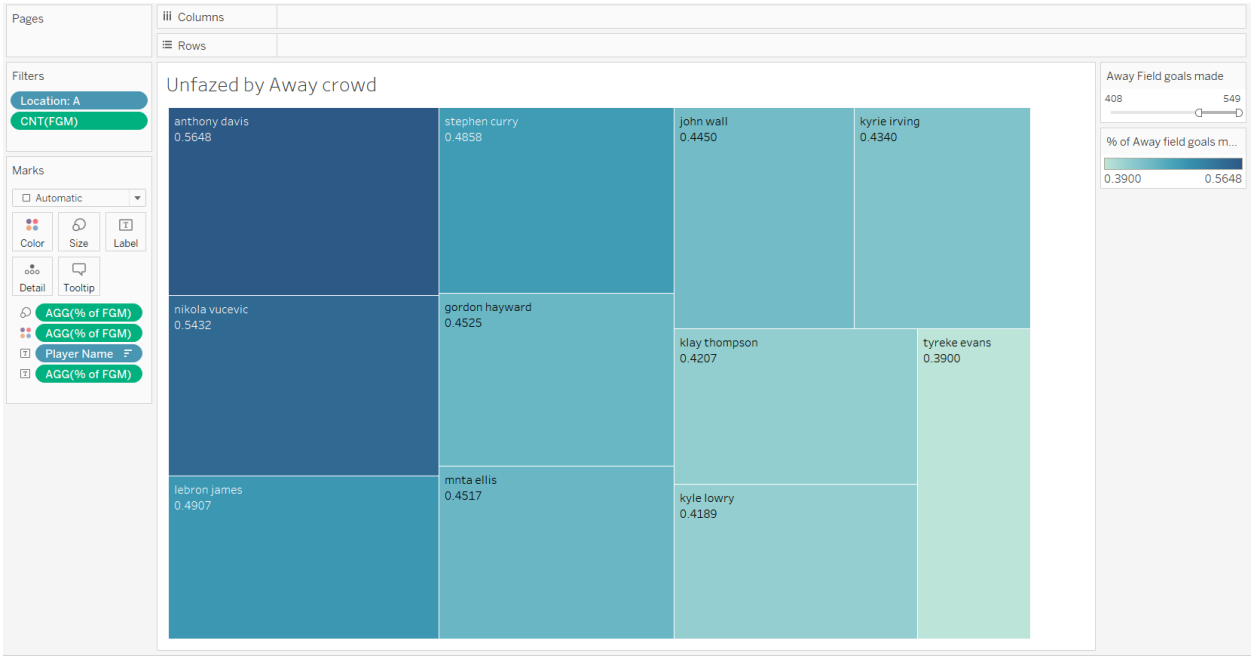
Going by the visualizations for top clutch players and ball hogs we can confirm that players can indeed make an impact without holding onto the ball for a long time.

- Should the NBA move to a more three pointer oriented style of play?



Although the league is moving more towards a three pointer jump shooting approach the visualizations for the best finishers in the paint still suggests that by playing closer to the basket there is higher chance of making the shot. This can be seen from the high field goal percentage for such shots attempted.

- Does playing away have an impact on top NBA player's performance?



Based on the field goal aggregate percentage we can confirm that playing away has not been detrimental to a top in the league NBA professional's performance.