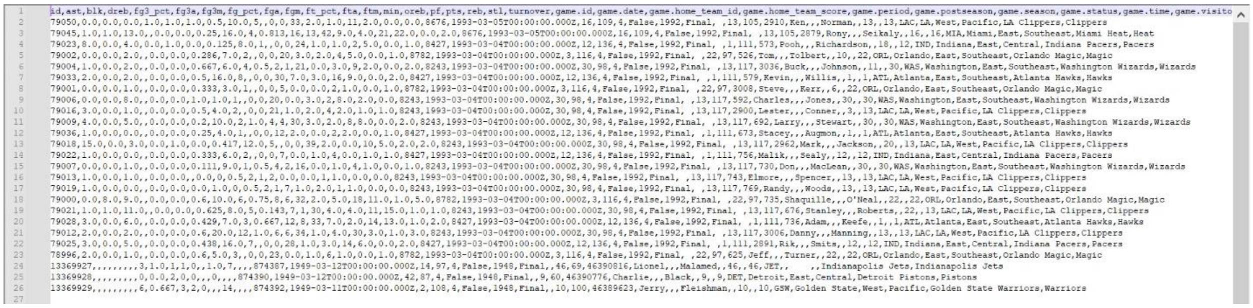
<u>Data Science 1</u>

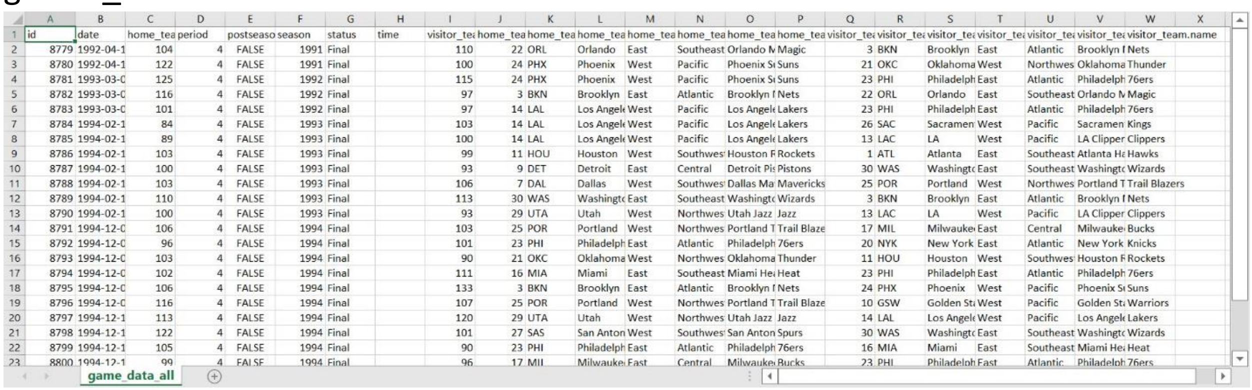
Screenshots:

stats_data



```
id,ast,blk,dreb,fg3_pct,fg3a,fg3m,fg_pct,fga,fgm,ft_pct,fta,ftm,min,oreb,pf,pts,reb,stl,turnover,game.id,game.date,game.home_team_id,game.home_team_score,game.period,game.postseason,game.season,game.status,game.time,game.visito
79050,0.0,0.0,0.0,1.0,1.0,1.0,0.5,10.0,5,,0,0,33,2.0,1.0,11,2.0,0.0,0.0,8676,1993-03-05T00:00:00.000Z,16,109,4,False,1992,Final, ,13,105,2910,Ken,,,Norman,,13,,13,LAC,LA,West,Pacific,LA Clippers,Clippers
79045,1.0,1.0,13.0,,0.0,0.0,0.25,16.0,4,0.813,16,13,42,9.0,4.0,21,22.0,0.0,2.0,8676,1993-03-05T00:00:00.000Z,16,109,4,False,1992,Final, ,13,105,2879,Rony,,,Seikaly,,16,,16,MIA,Miami,East,Southeast,Miami Heat,Heat
79023,8.0,0.0,4.0,0.0,1.0,0.0,0.125,8.0,1,,0,0,24,1.0,1.0,2,5.0,0.0,1.0,8427,1993-03-04T00:00:00.000Z,12,136,4,False,1992,Final, ,1,111,573,Pooh,,,Richardson,,18,,12,IND,Indiana,East,Central,Indiana Pacers,Pacers
79002,0.0,0.0,2.0,,0.0,0.0,0.286,7.0,2,,0,0,20,3.0,2.0,4,5.0,0.0,1.0,8782,1993-03-04T00:00:00.000Z,3,116,4,False,1992,Final, ,22,97,526,Tom,,,Tolbert,,10,,22,ORL,Orlando,East,Southeast,Orlando Magic,Magic
79004,1.0,0.0,2.0,,0.0,0.0,0.667,6.0,4,0.5,2,1,21,0.0,3.0,9,2.0,0.0,2.0,8243,1993-03-04T00:00:00.000Z,30,98,4,False,1992,Final, ,13,117,3036,Buck,,,Johnson,,11,,30,WAS,Washington,East,Southeast,Washington Wizards,Wizards
79033,2.0,0.0,2.0,,0.0,0.0,0.5,16.0,8,,0,0,30,7.0,3.0,16,9.0,0.0,2.0,8427,1993-03-04T00:00:00.000Z,12,136,4,False,1992,Final, ,1,111,579,Kevin,,,Willis,,1,,1,ATL,Atlanta,East,Southeast,Atlanta Hawks,Hawks
79001,0.0,0.0,1.0,,0.0,0.0,0.333,3.0,1,,0,0,5,0.0,0.0,2,1.0,0.0,1.0,8782,1993-03-04T00:00:00.000Z,3,116,4,False,1992,Final, ,22,97,3008,Steve,,,Kerr,,6,,22,ORL,Orlando,East,Southeast,Orlando Magic,Magic
79006,0.0,0.0,8.0,,0.0,0.0,1.0,1.0,1,,0,0,20,0.0,3.0,2,8.0,2.0,0.0,8243,1993-03-04T00:00:00.000Z,30,98,4,False,1992,Final, ,13,117,592,Charles,,,Jones,,30,,30,WAS,Washington,East,Southeast,Washington Wizards,Wizards
79016,3.0,0.0,1.0,,0.0,0.0,0.5,4.0,2,,0,0,21,1.0,2.0,4,2.0,1.0,1.0,8243,1993-03-04T00:00:00.000Z,30,98,4,False,1992,Final, ,13,117,2900,Lester,,,Conner,,3,,13,LAC,LA,West,Pacific,LA Clippers,Clippers
79009,4.0,0.0,5.0,,0.0,0.0,0.2,10.0,2,1.0,4,4,30,3.0,2.0,8,8.0,0.0,2.0,8243,1993-03-04T00:00:00.000Z,30,98,4,False,1992,Final, ,13,117,692,Larry,,,Stewart,,30,,30,WAS,Washington,East,Southeast,Washington Wizards,Wizards
79036,1.0,0.0,0.0,,0.0,0.0,0.25,4.0,1,,0,0,12,2.0,0.0,2,2.0,0.0,1.0,8427,1993-03-04T00:00:00.000Z,12,136,4,False,1992,Final, ,1,111,673,Stacey,,,Augmon,,1,,1,ATL,Atlanta,East,Southeast,Atlanta Hawks,Hawks
79018,15.0,0.0,3.0,0.0,1.0,0.0,0.417,12.0,5,,0,0,39,2.0,0.0,10,5.0,2.0,2.0,8243,1993-03-04T00:00:00.000Z,30,98,4,False,1992,Final, ,13,117,2962,Mark,,,Jackson,,20,,13,LAC,LA,West,Pacific,LA Clippers,Clippers
79022,1.0,0.0,0.0,,0.0,0.0,0.333,6.0,2,,0,0,7,0.0,1.0,4,0.0,1.0,1.0,8427,1993-03-04T00:00:00.000Z,12,136,4,False,1992,Final, ,1,111,756,Malik,,,Sealy,,12,,12,IND,Indiana,East,Central,Indiana Pacers,Pacers
79007,0.0,0.0,1.0,,0.0,0.0,0.111,9.0,1,0.5,4,2,16,0.0,1.0,4,1.0,0.0,1.0,8243,1993-03-04T00:00:00.000Z,30,98,4,False,1992,Final, ,13,117,730,Don,,,MacLean,,30,,30,WAS,Washington,East,Southeast,Washington Wizards,Wizards
79013,1.0,0.0,1.0,,0.0,0.0,,0.0,0,0.5,2,1,2,0.0,0.0,1,1.0,0.0,0.0,8243,1993-03-04T00:00:00.000Z,30,98,4,False,1992,Final, ,13,117,743,Elmore,,,Spencer,,13,,13,LAC,LA,West,Pacific,LA Clippers,Clippers
79019,1.0,0.0,0.0,,0.0,0.0,0.0,1.0,0,0.5,2,1,7,1.0,2.0,1,1.0,0.0,0.0,8243,1993-03-04T00:00:00.000Z,30,98,4,False,1992,Final, ,13,117,769,Randy,,,Woods,,13,,13,LAC,LA,West,Pacific,LA Clippers,Clippers
79000,0.0,8.0,9.0,,0.0,0.0,0.6,10.0,6,0.75,8,6,32,2.0,5.0,18,11.0,1.0,5.0,8782,1993-03-04T00:00:00.000Z,3,116,4,False,1992,Final, ,22,97,735,Shaquille,,,O'Neal,,22,,22,ORL,Orlando,East,Southeast,Orlando Magic,Magic
79021,1.0,1.0,11.0,,0.0,0.0,0.625,8.0,5,0.143,7,1,30,4.0,4.0,11,15.0,1.0,1.0,8243,1993-03-04T00:00:00.000Z,30,98,4,False,1992,Final, ,13,117,676,Stanley,,,Roberts,,22,,13,LAC,LA,West,Pacific,LA Clippers,Clippers
79028,3.0,0.0,6.0,,0.0,0.0,0.429,7.0,3,0.667,12,8,33,7.0,2.0,14,13.0,1.0,2.0,8427,1993-03-04T00:00:00.000Z,12,136,4,False,1992,Final, ,1,111,736,Adam,,,Keefe,,1,,1,ATL,Atlanta,East,Southeast,Atlanta Hawks,Hawks
79012,2.0,0.0,2.0,,0.0,0.0,0.6,20.0,12,1.0,6,6,34,1.0,4.0,30,3.0,1.0,3.0,8243,1993-03-04T00:00:00.000Z,30,98,4,False,1992,Final, ,13,117,3006,Danny,,,Manning,,13,,13,LAC,LA,West,Pacific,LA Clippers,Clippers
79025,3.0,0.0,5.0,,0.0,0.0,0.438,16.0,7,,0,0,28,1.0,3.0,14,6.0,0.0,2.0,8427,1993-03-04T00:00:00.000Z,12,136,4,False,1992,Final, ,1,111,2891,Rik,,,Smits,,12,,12,IND,Indiana,East,Central,Indiana Pacers,Pacers
78996,2.0,0.0,1.0,,0.0,0.0,0.6,5.0,3,,0,0,23,0.0,1.0,6,1.0,0.0,1.0,8782,1993-03-04T00:00:00.000Z,3,116,4,False,1992,Final, ,22,97,625,Jeff,,,Turner,,22,,22,ORL,Orlando,East,Southeast,Orlando Magic,Magic
13369927,,,,,,,,,3,1.0,1,1,0,,1.0,7,,,,874387,1949-03-12T00:00:00.000Z,14,97,4,False,1948,Final,,46,69,46390816,Lionel,,,Malamed,,46,,46,JET,,    ,,Indianapolis Jets,Indianapolis Jets
13369928,,,,,,,,,0,0.0,2,0,0,,,0,,,,874390,1949-03-12T00:00:00.000Z,42,87,4,False,1948,Final,,9,60,46390776,Charlie,,,Black,,9,,9,DET,Detroit,East,Central,Detroit Pistons,Pistons
13369929,,,,,,,,,6,0.667,3,2,0,,,14,,,,874392,1949-03-11T00:00:00.000Z,2,108,4,False,1948,Final,,10,100,46389623,Jerry,,,Fleishman,,10,,10,GSW,Golden State,West,Pacific,Golden State Warriors,Warriors
```

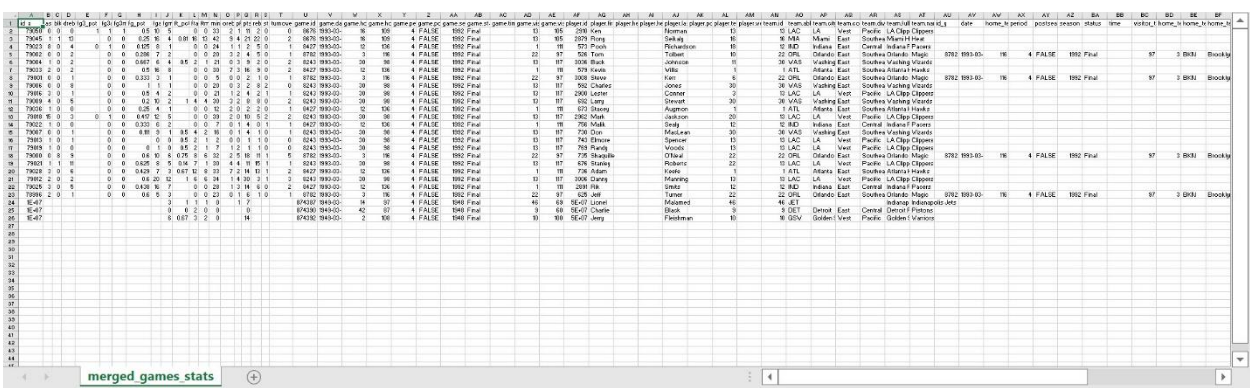
games_data



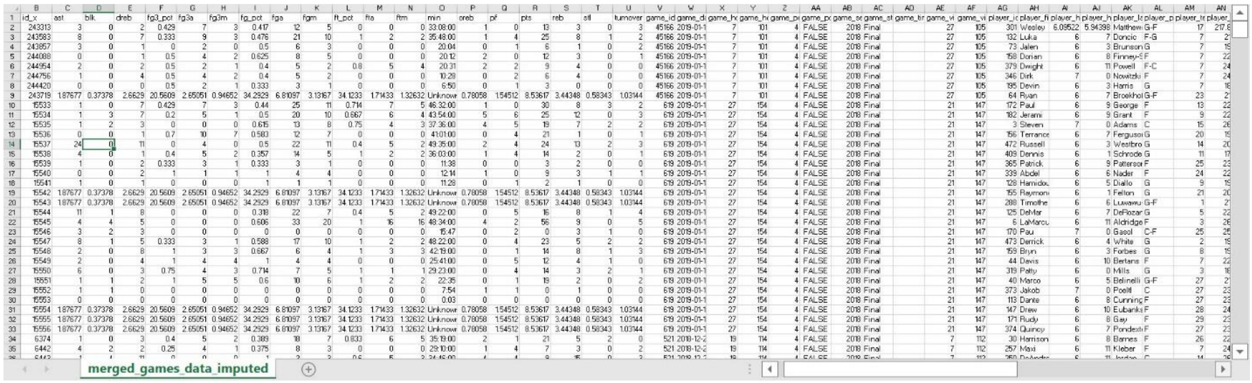
id	date	home_tea	period	postseaso	season	status	time	visitor_te	home_tea	home_tea	home_tea	home_tea	home_tea	home_tea	home_tea	visitor_te	visitor_te	visitor_te	visitor_te	visitor_te	visitor_te	visitor_team.name
8779	1992-04-1	104	4	FALSE	1991	Final		110	22	ORL	Orlando	East	Southeast	Orlando M	Magic	3	BKN	Brooklyn	East	Atlantic	Brooklyn I	Nets
8780	1992-04-1	122	4	FALSE	1991	Final		100	24	PHX	Phoenix	West	Pacific	Phoenix S	Suns	21	OKC	Oklahoma	West	Northwes	Oklahoma	Thunder
8781	1993-03-0	125	4	FALSE	1992	Final		115	24	PHX	Phoenix	West	Pacific	Phoenix S	Suns	23	PHI	Philadelph	East	Atlantic	Philadelph	76ers
8782	1993-03-0	116	4	FALSE	1992	Final		97	3	BKN	Brooklyn	East	Atlantic	Brooklyn I	Nets	22	ORL	Orlando	East	Southeast	Orlando M	Magic
8783	1993-03-0	101	4	FALSE	1992	Final		97	14	LAL	Los Angel	West	Pacific	Los Angel	Lakers	23	PHI	Philadelph	East	Atlantic	Philadelph	76ers
8784	1994-02-1	84	4	FALSE	1993	Final		103	14	LAL	Los Angel	West	Pacific	Los Angel	Lakers	26	SAC	Sacramen	West	Pacific	Sacramen	Kings
8785	1994-02-1	89	4	FALSE	1993	Final		100	14	LAL	Los Angel	West	Pacific	Los Angel	Lakers	13	LAC	LA	West	Pacific	LA Clipper	Clippers
8786	1994-02-1	103	4	FALSE	1993	Final		99	11	HOU	Houston	West	Southwes	Houston R	Rockets	1	ATL	Atlanta	East	Southeast	Atlanta Ha	Hawks
8787	1994-02-1	100	4	FALSE	1993	Final		93	9	DET	Detroit	East	Central	Detroit Pis	Pistons	30	WAS	Washingtc	East	Southeast	Washingtc	Wizards
8788	1994-02-1	103	4	FALSE	1993	Final		106	7	DAL	Dallas	West	Southwes	Dallas Ma	Mavericks	25	POR	Portland	West	Northwes	Portland T	Trail Blazers
8789	1994-02-1	110	4	FALSE	1993	Final		113	30	WAS	Washingtc	East	Southeast	Washingtc	Wizards	3	BKN	Brooklyn	East	Atlantic	Brooklyn I	Nets
8790	1994-02-1	100	4	FALSE	1993	Final		93	29	UTA	Utah	West	Northwes	Utah Jazz	Jazz	13	LAC	LA	West	Pacific	LA Clipper	Clippers
8791	1994-12-0	106	4	FALSE	1994	Final		103	25	POR	Portland	West	Northwes	Portland T	Trail Blaze	17	MIL	Milwauke	East	Central	Milwauke	Bucks
8792	1994-12-0	96	4	FALSE	1994	Final		101	23	PHI	Philadelph	East	Atlantic	Philadelph	76ers	20	NYK	New York	East	Atlantic	New York	Knicks
8793	1994-12-0	103	4	FALSE	1994	Final		90	21	OKC	Oklahoma	West	Northwes	Oklahoma	Thunder	11	HOU	Houston	West	Southwes	Houston R	Rockets
8794	1994-12-0	102	4	FALSE	1994	Final		111	16	MIA	Miami	East	Southeast	Miami He	Heat	23	PHI	Philadelph	East	Atlantic	Philadelph	76ers
8795	1994-12-0	106	4	FALSE	1994	Final		133	3	BKN	Brooklyn	East	Atlantic	Brooklyn I	Nets	24	PHX	Phoenix	West	Pacific	Phoenix S	Suns
8796	1994-12-0	116	4	FALSE	1994	Final		107	25	POR	Portland	West	Northwes	Portland T	Trail Blaze	10	GSW	Golden St	West	Pacific	Golden St	Warriors
8797	1994-12-1	113	4	FALSE	1994	Final		120	29	UTA	Utah	West	Northwes	Utah Jazz	Jazz	14	LAL	Los Angel	West	Pacific	Los Angel	Lakers
8798	1994-12-1	122	4	FALSE	1994	Final		101	27	SAS	San Anton	West	Southwes	San Anton	Spurs	30	WAS	Washingtc	East	Southeast	Washingtc	Wizards
8799	1994-12-1	105	4	FALSE	1994	Final		90	23	PHI	Philadelph	East	Atlantic	Philadelph	76ers	16	MIA	Miami	East	Southeast	Miami He	Heat
8800	1994-12-1	99	4	FALSE	1994	Final		96	17	MIL	Milwauke	East	Central	Milwauke	Bucks	23	PHI	Philadelph	East	Atlantic	Philadelph	76ers

game_data_all

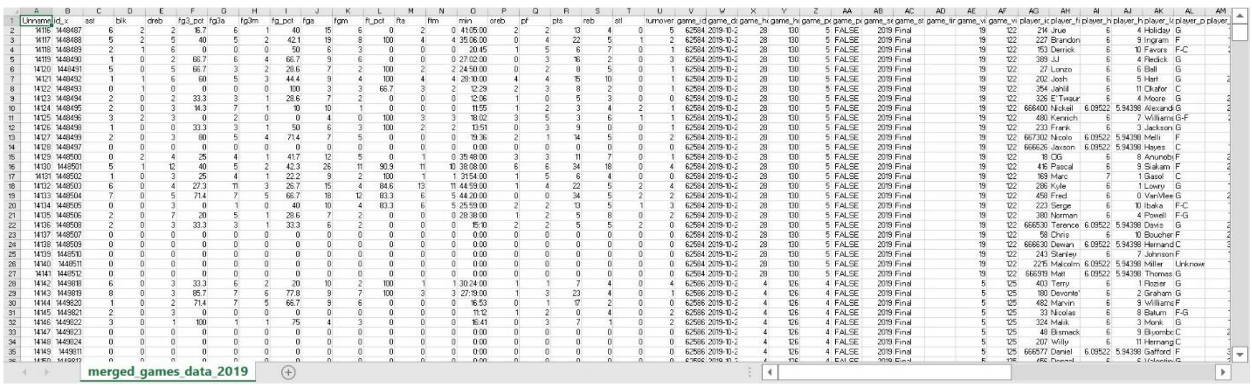
merged_on_8782



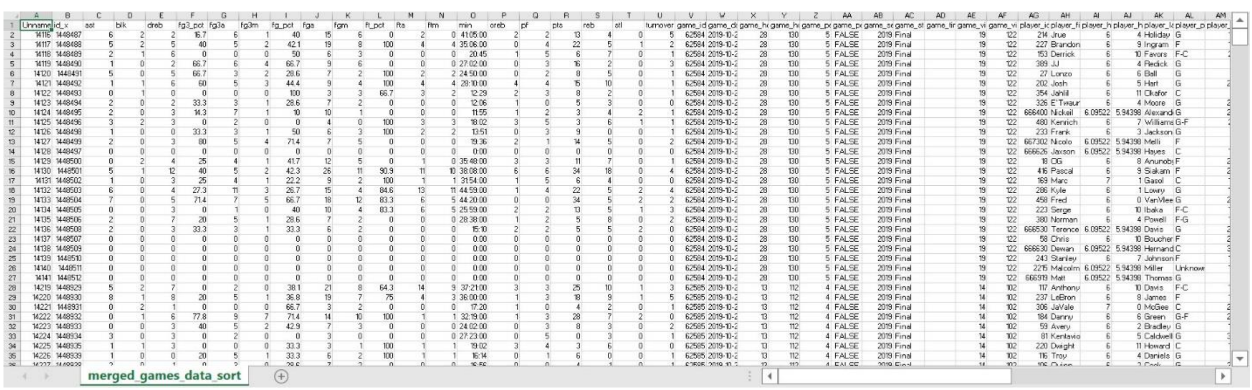
merged_imputed



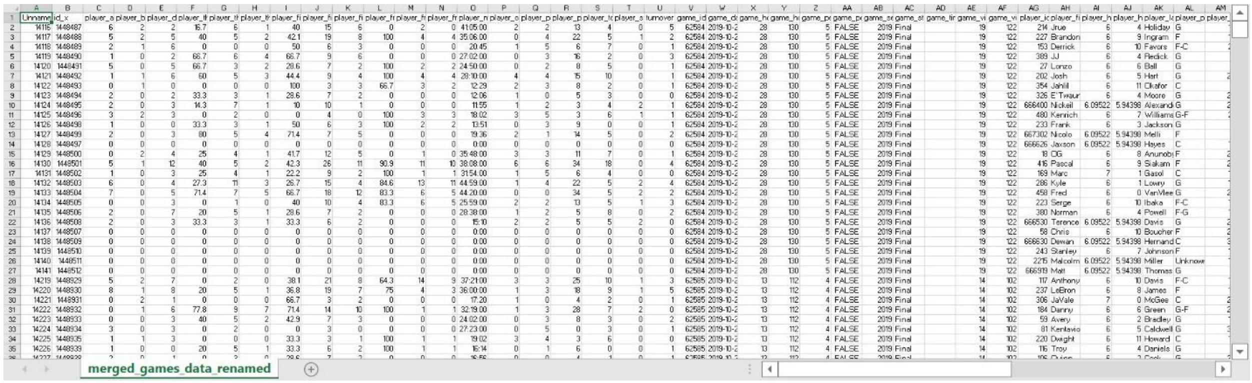
merged_2019



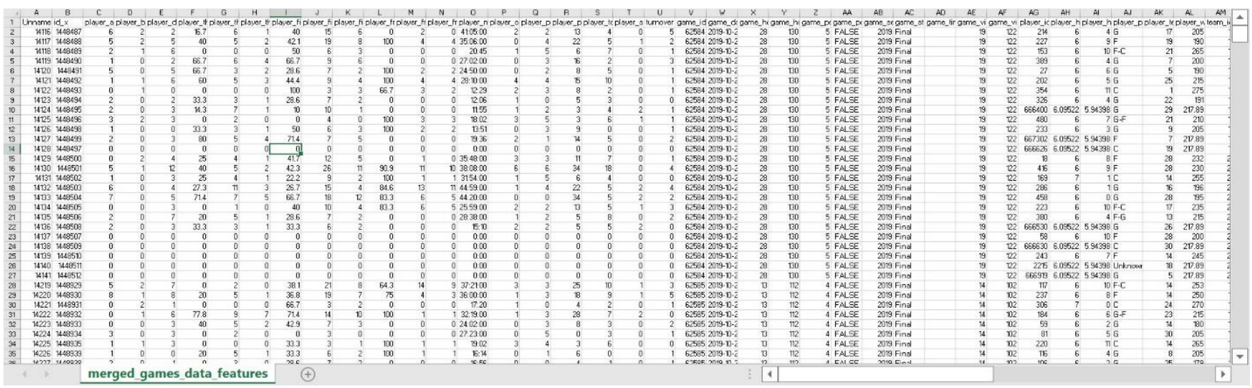
merged_sorted



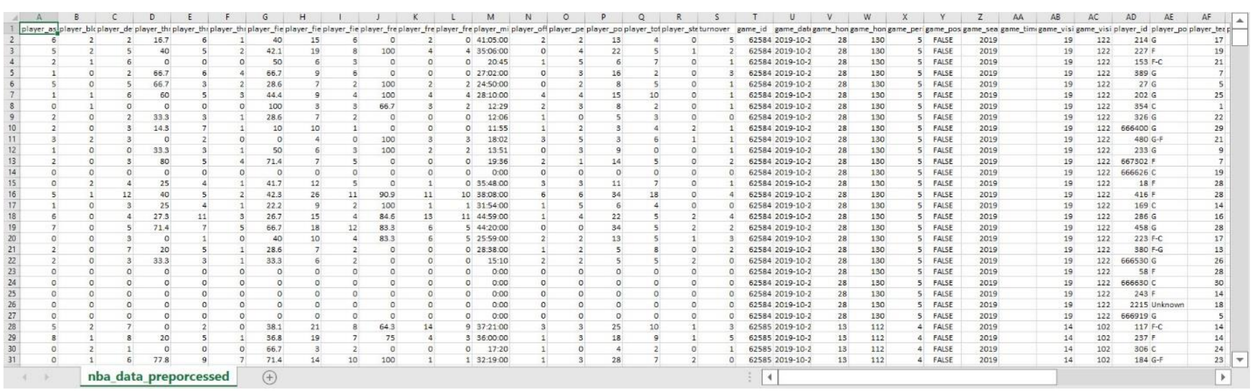
merged_renamed



merged_features



nba_preprocessed



Headers in final data frame:

player_assists, player_blocks, player_defensive_rebounds, player_three_point_pct, player_three_point_attempts, player_three_point_made, player_field_goal_pct, player_field_goal_attempts, player_field_goal_made, player_free_throw_pct, player_free_throw_attempts, player_free_throw_made, player_minutes_played, player_offensive_rebounds, player_personal_fouls, player_points, player_total_rebounds, player_steals, turnover, game_id, game_date, game_home_team_id, game_home_team_score, game_period, game_postseason, game_season, game_time, game_visitor_team_id, game_visitor_team_score, player_id, player_position, player_team_id, player_weight_pounds, team_id, team_division, team_full_name, home_team_conference, home_team_division, home_team_full_name, visitor_team_conference, visitor_team_division, visitor_team_full_name, team_won, abs_score_difference, player_full_name, player_height_cm, home_player_stats, player_efficiency, player_three_to_throw_attempt_ratio, player_three_to_throw_made_ratio

List of new features:

player_efficiency: This feature represents the player's efficiency in the game. It is calculated by summing the player's points, total rebounds, assists, steals, and blocks, and then subtracting the missed field goal attempts, missed free throw attempts, and turnovers. A higher value indicates a more efficient performance by the player.

player_three_to_throw_attempt_ratio: This feature represents the ratio of three-point attempts to total field goal attempts by the player. It is calculated by dividing the number of three-point attempts by the number of total field goal attempts. It provides insight into the player's preference for shooting three-pointers compared to other types of field goals.

player_three_to_throw_made_ratio: This feature represents the ratio of three-point made shots to total field goals made by the player. It is calculated by dividing the number of three-point made shots by the number of total field goals made. It gives an indication of the player's accuracy and success rate in making three-point shots compared to other types of field goals.

These features provide additional insights into the player's overall performance, shooting tendencies, and efficiency on the court.

Answers to questions:

The team with the most wins is: ('Los Angeles Lakers', 46)

The team with the least wins is: ('Charlotte Hornets', 31)

Total LA Lakers points: 10297

Total Charlotte Hornets points: 6687

a. Supervised learning refers to the type of machine learning where the model is trained on labeled data, where the input data is paired with corresponding target labels. The goal is to learn a mapping function that can predict the labels for new, unseen data. Some use cases for supervised learning include email spam classification, sentiment analysis, and image recognition.

Unsupervised learning, on the other hand, involves training a model on unlabeled data without any predefined target labels. The model learns patterns, structures, or relationships within the data without specific guidance. Use cases for unsupervised learning include clustering similar documents, customer segmentation, and anomaly detection.

b. Structured data refers to data that is organized and well-defined, typically stored in databases or structured file formats. It has a fixed schema and follows a consistent format, making it easy to analyze using traditional database operations. Use cases for structured data include financial transactions, sales records, and sensor data from IoT devices.

Unstructured data, on the other hand, does not adhere to a specific schema or format. It includes text documents, images, audio files, videos, social media posts, etc. Analyzing unstructured data requires specialized techniques such as natural language processing, computer vision, or audio processing. Use cases for unstructured data include social media sentiment analysis, image recognition, and speech-to-text conversion.

c. The Cross Industry Standard Process for Data Mining (CRISP-DM) is a widely used data mining process model. The main steps of CRISP-DM are:

Business Understanding: Understand the project objectives, requirements, and constraints from a business perspective. Define the data mining problem and set goals.

Data Understanding: Gather and explore the available data to gain insights into its structure, content, and quality. Identify any data issues or limitations.

Data Preparation: Select, clean, and transform the data to create a suitable dataset for modeling. This involves tasks like handling missing values, encoding categorical variables, and normalizing data.

Modeling: Select appropriate modeling techniques and build a predictive or descriptive model based on the prepared dataset. This step includes model training, validation, and evaluation.

Evaluation: Assess the quality and effectiveness of the developed model. Evaluate its performance against the project objectives and requirements.

Deployment: Deploy the model into a production environment, integrating it with existing systems or processes. Monitor the model's performance and make necessary adjustments.

Some use cases include customer churn prediction, fraud detection, and market basket analysis.

d. A feature vector is a representation of an object or data point in a machine learning model. It is a numerical representation that captures the relevant characteristics or features of the data. Feature vectors are used as input to machine learning algorithms for training and making predictions.

In general, feature vectors can be used to represent various types of data, such as text documents (bag-of-words representation), images (pixel values or extracted features), or audio signals (MFCC coefficients). In the preprocessed data, feature vectors can be derived from the extracted features of the basketball game data, such as player statistics, team performance metrics, or game-related attributes.

Some use cases of feature vectors include sentiment analysis of text, image classification, and speech recognition.

e. Feature engineering is the process of creating new features or transforming existing features to improve the performance of machine learning models. It involves selecting, combining, or deriving features that capture the most relevant information from the data.

In general, feature engineering aims to enhance the predictive power of the model, reduce overfitting, and improve generalization. It can involve techniques like feature scaling, one-hot encoding, feature extraction, or feature selection. In the preprocessed data, feature engineering could include creating derived features like player efficiency, three-point-to-field-goal attempt ratio, or any other