| Date | Duration | Type |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

**Reflection**

Main goals in this time period?

Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help?

| Description of work | Challenges/Next Steps |
| --- | --- |
| | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

| Date | Duration | Type |
|---|---|---|
| 8/28/2024 | 1 hour | Other |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

**Reflection**

Main goals in this time period?

Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help?

| Description of work | Challenges/Next Steps |
|---|---|
| Orientation | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

| Date | Duration | Type |
|---|---|---|
| ######## | 3 hours | Research |
| ######## | 4 hours | Coding |
| ######## | 1 hour | Office Hours |
| | | |
| | | |
| | | |
| | | |
| | | |

**Reflection**

| | |
|---|---|
| Main goals in this time period? | Getting started and seeing where I could get my data from. |
| Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help? | This API isn't made by the NBA and the API they do have isn't pu |

| Description of work | Challenges/Next Steps |
|---|---|
| Reviewed NBA API documentation, exploring endpoints for player stats | Understanding how to access complete datasets efficiently. |
| Tested API endpoints to retrieve player data. | Timeout errors and difficulty in retrieving stats for all players at once. |
| Deadline reminders and project ideas | |
| | |
| | |
| | |
| | |
| | |

ublicly accessible

| Date | Duration | Type |
|---|---|---|
| 9/17/2024 | 5 hours | Debugging |
| 9/19/24 | 4 hours | Research |
| 9/14/2024 | 2 hours | Debugging |
| | | |
| | | |
| | | |
| | | |
| | | |

**Reflection**

Main goals in this time period? Getting data from the API

Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help?

Starting to have problems with reading documentation and rate limit

| Description of work | Challenges/Next Steps |
| --- | --- |
| Wrote Python scripts to query NBA API for individual player stats by IDs. | Struggling with incomplete data and redundant API calls. |
| Searched for alternative approaches, such as using multiple endpoints. | Balancing exploration with progress on cleaning data for analysis. |
| Investigated API rate limits and strategies for handling timeouts. | Experimenting with delays to avoid overwhelming the API. |

| | |
| --- | --- |
| | |
| | |
| | |
| | |

ing

| Date | Duration | Type |
|---|---|---|
| 9/24/2024 | 4 hours | Coding |
| 9/26/2024 | 5 hours | Debugging |
| 9/21/24 | 3 hours | Documentation |

| | | |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |

**Reflection**

| | |
|---|---|
| Main goals in this time period? | Format data that I've gotten from the API so far |
| Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help? | API is proving hard to work with so I'm thinking of scraping data. |

| Description of work | Challenges/Next Steps |
|---|---|
| Used API data to start creating initial player datasets. | Duplicated player entries due to mismatched IDs; refined queries. |
| Tried to consolidate multiple API responses into a single usable dataset. | Slow queries due to rate limiting; incomplete results for low-minute players. |
| Logged challenges with API usage, noting key limitations. | Deciding whether to continue with the NBA API or switch to scraping. |

| | |
|---|---|
| | |
| | |
| | |
| | |

| Date | Duration | Type |
|---|---|---|
| 9/28/2024 | 2 hours | Reflection |
| 10/2/2024 | 6 hours | Research |
| 10/4/2024 | 4 hours | Coding |
| 10/7/2024 | 1 hour | Office Hours |
| | | |
| | | |
| | | |
| | | |
| | | |

**Reflection**

| | |
|---|---|
| Main goals in this time period? | Continuing to grab my data but scraping seems to be the best ch |
| Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help? | Coping with starting from square one as I've already spent time |

| Description of work | Challenges/Next Steps |
|---|---|
| Evaluated progress and weighed the pros/cons of continuing with the API. | Exploring alternatives, including web scraping. |
| Tested basic web scraping techniques using BeautifulSoup | |
| Began scraping Basketball Reference for college season averages. | Addressing cases with missing player data. |

Deadline reminders and project ideas

| | |
|---|---|
| | |
| | |
| | |
| | |

ıoice.

on the API.

**Reflection**

| | |
|---|---|
| Main goals in this time period? | Building scripts to automate scrapes, then cleaning data. |
| Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help? | Scraping is way easier than using the API. But I fear that I need to think about how I'm going to account for data outliers and why they occur. I should probably take players that average more than 15 min and 20 games per season |

| Date | Duration | Type | Description of work |
|---|---|---|---|
| 10/8/2024 | 3 hours | Data Cleaning | Cleaned initial scrape results |
| ######## | 5 hours | Coding | Expanded scraping to include career averages from Basketball Reference. |
| ######## | 4 hours | Data Cleaning | Processed scraped data |
| ######## | 1 hour | Office Hours | Deadline reminders and project ideas, NYT internship |
| | | | |
| | | | |
| | | | |
| | | | |

| Challenges/Next Steps |
| --- |
| Handling rate limits and missing player data during scraping. |
| |
| Addressing discrepancies between datasets |
| |
| |
| |
| |
| |

| Date |
| --- |
| |
| ######## |
| ######## |
| ######## |
| |
| |
| |
| |
| |

**Reflection**

| | |
| --- | --- |
| Main goals in this time period? | Cleaning and formatting |
| Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help? | I have so many ideas for how to analyze this that I want to format the data in so many ways. I need to consolidate my analysis and keep the project simple. I'm doing the Amazon PRFAQ process |

| Duration | Type | Description of work | Challenges/Next Steps |
|---|---|---|---|
| 6 hours | Coding | Completed scripts to scrape NCAA stats for the top five conferences (2003–24). | Resolving inconsistencies in team and player data across years. Thinking of how to format data, multiple ways perhaps |
| 3 hours | Data Cleaning | Cleaned and merged NCAA datasets with scraped NBA career stats. | Matching player data across datasets while avoiding duplication. |
| 4 hours | Research | BC Alumni from Amazon visited class and explained some Amazon processes. | Thinking of implementing the PRFAQ process to my project. |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

| Date | Duration |
|---|---|
| ######## | 5 hours |
| ######## | 4 hours |
| ######## | 3 hours |
| | |
| | |
| | |
| | |
| | |

**Reflection**

| | |
|---|---|
| Main goals in this time period? | Cleaning |
| Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help? | So many nuances that I want to worry about in terms of which segment of college players that I want to |

| Type | Description of work | Challenges/Next Steps |
|---|---|---|
| EDA Preparation | Finalized cleaning steps and prepared datasets for exploratory analysis. | Ensuring uniformity across datasets before analysis begins. |
| Research/EDA | Explored trends in NCAA player stats across conferences. | Identifying patterns and starting to define categories for players. |
| Visualization | Created histograms and boxplots to visualize stat distributions by conference. | |
| | | |
| | | |
| | | |
| | | |
| | | |

| Date | Duration | Type |
|---|---|---|
| ######## | 4 hours | EDA |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

**Reflection**

| | |
|---|---|
| Main goals in this time period? | Starting on EDA |
| Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help? | EDA is my favorite part! |

| Description of work | Challenges/Next Steps |
| --- | --- |
| Compared conference stats | |
| | |
| | |
| | |
| | |
| | |
| | |
| | |

**Reflection**

| | |
|---|---|
| Main goals in this time period? | Getting back into the swing of things by cleaning data and getting it in be |
| Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help? | I was confused with what the issue was when it came to building these tables that said who from college made it to the NBA. After a while, I saw that names were formatted differently between the college and NBA data sets. |

| Date | Duration | Type | Description of work |
|---|---|---|---|
| ######## | 5 hours | Data Cleaning | Standardized NCAA and NBA datasets for consistent feature engineering. |
| ######## | 5 hours | EDA | Conducted deeper analysis into trends for impact scores across player groups. |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

etter shape for EDA. I had some issues with EDA last time due to some formatting

| Challenges/Next Steps |
| --- |
| Finalizing adjustments for alignment issues with player names. |
| |
| |
| |
| |
| |
| |
| |

issues between datasets.

**Reflection**

| | |
|---|---|
| Main goals in this time period? | Cleaning and EDA |
| Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help? | Having a hard time focusing on this project after the election. This election directly affected me and I went to office hours for some words of wisdom. Need a story of somebody that's been in my position and made it out fine. |

| Date | Duration | Type | Description of work |
|---|---|---|---|
| ######## | 5 hours | Data Cleaning | Addressed remaining issues with aligning NCAA and NBA data. |
| ######## | 4 hours | EDA | Explored correlations between NCAA stats and getting into NBA. |
| ######## | 3 hours | Visualization | Refined visualizations showing stat trends for NCAA players entering the NBA. |
| ######## | 1 hour | Office Hours | Talking about future as a professional affected by election |
| | | | |
| | | | |
| | | | |
| | | | |

| Challenges/Next Steps |
| --- |
| Standardizing player names and filtering data for key features |
| Identifying meaningful relationships to guide impact scoring. |
| Ensuring visuals effectively highlight patterns in player performance. |
| Not sure what to do |
| |
| |
| |
| |

**Reflection**

| | |
|---|---|
| Main goals in this time period? | Classifying NBA players into tiers. |
| Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help? | Having a hard time going back and forth between reformatting data and analyzing. I wanted to first categorize players using percentiles but maybe classifying using the data itself through k-means is a better idea? |

| Date | Duration | Type | Description of work |
|---|---|---|---|
| ######## | 5 hours | Research/EDA | Compared distributions of impact scores for players across NCAA conferences. |
| ######## | 4 hours | PRFAQ Documentation | Drafting my PRFAQ |
| ######## | 3 hours | Data Refinement | Verified consistency in stats across years and conferences. |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

| Challenges/Next Steps |
| --- |
| Handling skewed distributions caused by low-minute players. |
| |
| Preparing the dataset for future clustering and machine learning. |
| |
| |
| |
| |
| |

| Date | Duration |
|---|---|
| ######## | 4 hours |
| ######## | 5 hours |
| 12/1/2023 | 3 hours |
| | |
| | |
| | |
| | |
| | |

**Reflection**

| | |
|---|---|
| Main goals in this time period? | Analyzing and creating models |
| Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help? | Deadline is approaching and I'm about 65-70% done. Need to focus on having a finished product instead of being fancy. |

| Type | Description of work | Challenges/Next Steps |
|---|---|---|
| Analysis | Computed percentile-based impact scores and explored clustering readiness. | Deciding whether to move forward with K-Means or a simpler approach. |
| Coding/Research | Finalized derived features for clustering inputs | Refining features to improve clustering results. |
| Visualization | Created initial visuals to compare clustering readiness across metrics. | Validating whether input features effectively separate player tiers. |
| | | |
| | | |
| | | |
| | | |
| | | |

**Reflection**

Main goals in this time period?

Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help?

Deciding on model selection and looking at how I can classify players into tiers.

So it looks like no matter what I do, k means will classify players that I don't want classified as elite into the elite tier. This makes sense since K-Means prioritizes variance minimization, not basketball context. So I'm going to pivot and use percentiles as my classifiers since percentiles create clear thresholds (e.g., Top 10% as elite) that better reflect basketball hierarchies, robust against skewed distributions or outliers, and it's probably easier to

| Date | Duration | Type | Description of work |
|---|---|---|---|
| 12/2/2024 | 2 hours | Machine Learning | Clustering based on impact score formula that I made |
| 12/3/2024 | 1 hour | Machine Learning | Reformatting data to include per 36 minute stats |
| 12/4/2024 | 2 hours | Machine Learning | Deciding between per 36 minute stats or raw career average. Even debating using pure percentile based classification method. |
| 5-Dec | 1 hour | Research | Using a different formula to determine impact score. |
| 6-Dec | 1 hour | Machine Learning | I'm getting the same issue and having trouble thinking about what to do |
| | | | |
| | | | |
| | | | |

| Challenges/Next Steps |
| --- |
| There's a problem because my clusters are containing players I don't want them to |
| Perhaps using per 36 minute stats will get rid of players that I don't want in certain clusters |
| Do I keep it simple or be fancy and let the data decide. Would using percentiles make my data more accurate? Seems like the clustering is more accurate on the 36 minute stats but there's still players I don't want in certain tiers |
| Deriving it from the PIE stat, but instead of accounting for games played, I won't |
| |
| |
| |
| |

**Reflection**

Main goals in this time period?

Main challenges during this phase? Were you able to meet the challenge, if so, what helped? It not, what could help?

Having a final product that works, creating slides and PRFAQ for better understanding for users. It's es

Really had trouble coping with the imperfections of my model. For some reason, I thought what I would make on the first attempt would be at least 70% accurate and be a ground breaker. But reflecting allowed me to be realistic and understand that if I was able to do that, NBA front offices probably wouldn't have a problem predicting which college players will excel. I'm proud of my work and will definitely improve on what I have beyond the semester.

| Date | Duration | Type | Description of work |
|---|---|---|---|
| 12/7/2024 | 3 hours | Machine Learning | Since I'm trying to predict the first 4 years of a college player, it makes sense to use NBA players' first four years instead of their entire career averages in my dataset. Time to reformat once again. |
| 12/8/2024 | 2 hours | Other | Implementing the PR FAQ process within my project so stakeholders and users will have a clearer idea of what I'm trying to make. |
| 12/9/2024 | 3 hours | Machine Learning | Decided on using random forest regressor and wrapping it around a multi output regressor to simultaneously predict multiple features. |
| ######## | 5 hours | Machine Learning | Doing resiudal analysis to see how well the predictions are turning out. Also using statistical methods to see how accurate my model is. Also creating the function that will do the predicting. |
| ######## | 2 Hours | Other | Creating slides |
| 12-Dec | 4 hours | Reflection & Documentation | The product works, it's just not as accurate as I want it to be. Of course it'll need more work. |
| | | | |
| | | | |

sentially a fancy Read me file.

| Challenges/Next Steps |
| --- |
| Focusing on players' first four years will get rid of plenty of players since it often takes more than 4 years for a player to break out. |
| |
| |
| |
| |
| |
| |
| |