Lara Weddige

Data Analytics Case Studies

Find me here

- Email
- LinkedIn
- GitHub
- Tableau

By Twitter, Fair use, https://en.wikipedia.org/w/index.php?curid=75492638

Table of contents

Analyzing global video game sales

GameCo

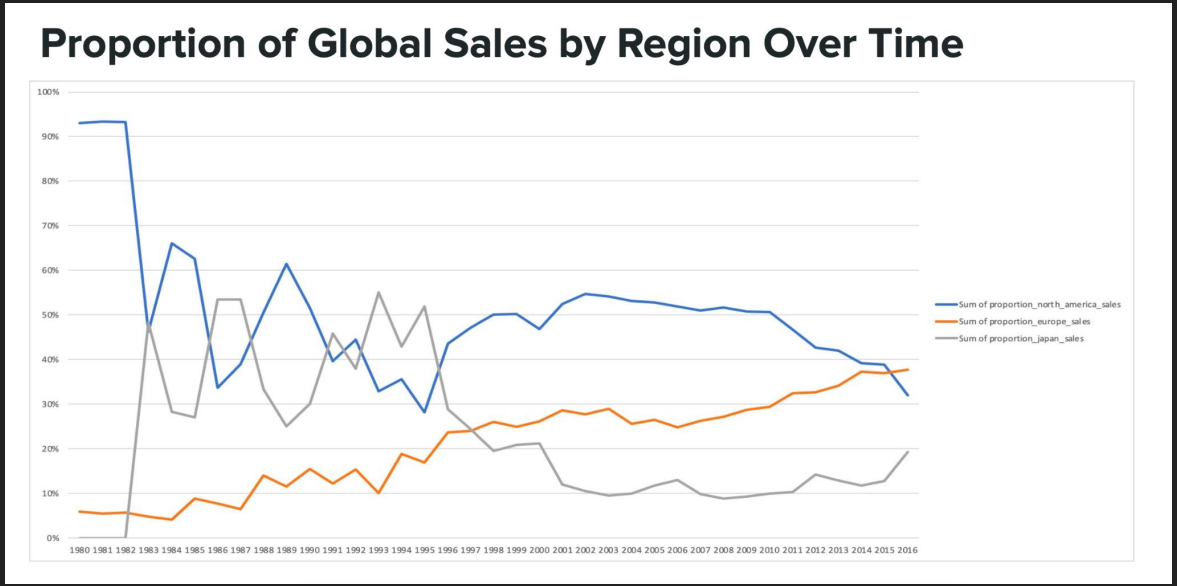
- Excel
- Grouping data
- Summarizing data
- Descriptive analysis
- Visualizing results in Excel
- Presenting results

Objectives, data, and general overview

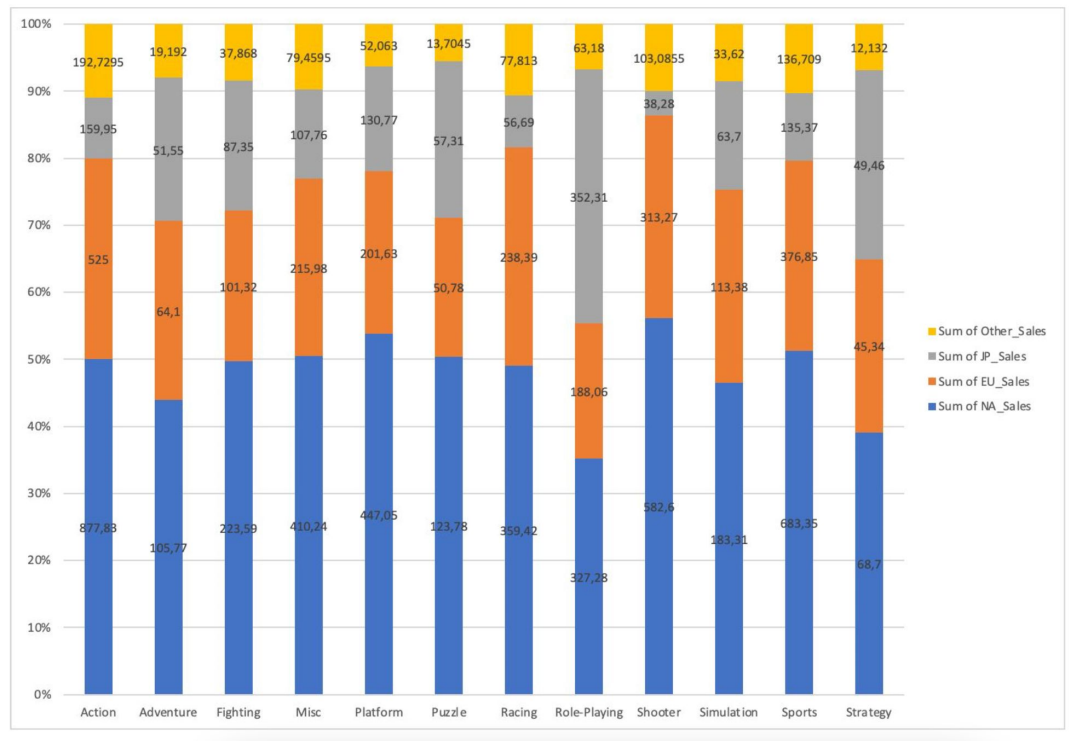
- **Objectives**:
  - Use data to inform the development of new games
  - Perform a descriptive analysis of a video game data set to foster a better understanding of how GameCo's new games might fare in the market
- **Data**:
  - Company data set from 1980 to 2016 that covers sales of video games (for games that sold more than 10,000 copies) spanning different platforms, genres, and publishing studios (data from VGChartz)
- **General overview**:
  - Challenging GameCo's current operating assumption of how video game sales behave across geographic regions, that sales for different areas have remained the same over time

Findings

- Swings in Japanese and North American sales making up the largest percentage of global sales from early 80s to mid 90s
- A gradual and steady rise in European sales up to its recent overtaking of North American sales in overall percentage
- Significant changes across the regions are opportunities for the marketing team
- This finding is counter to our previously held assumption that sales across regions have remained the same over time.



Findings



- Role-playing games are extremely popular in Japan
  - Much lower popularity in North American and Europe
- Shooter games are most popular in North America
  - Almost no sales in Japan
- Action and sports games account for the most sales overall
  - Very low demand in Japan
- Each region has their own preferences in terms of favorite genre and these differences are opportunities for the marketing team
- This finding underlines the need for localized marketing efforts.

Recommendations

Given that our initial assumption of sales remaining constant across different regions is not in line with what the data reveal:

- Redistribute marketing budget to maximize return on investment
  - Investing more in sales in Europe to encourage continued growth
  - Tapping into the recent upswing in Japanese sales
- Examine reasons behind currently falling North American sales

Given that the game genres enjoy varying levels of popularity across geographic regions:

- Distribute marketing budget to maximize return on investment
  - Focusing on the role-playing segment in Japan
  - Create campaigns about shooter games for North America, but not Japan
  - Highlight action and sports games
- Discover more information about other ways in which Japanese customers are different from North Americans or Europeans

Preparing for flu season in the US

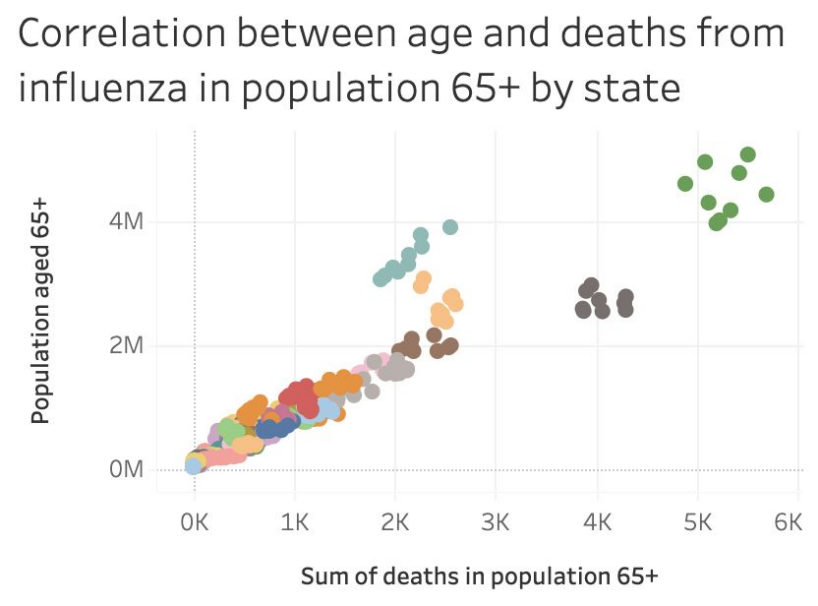
CDC + Census

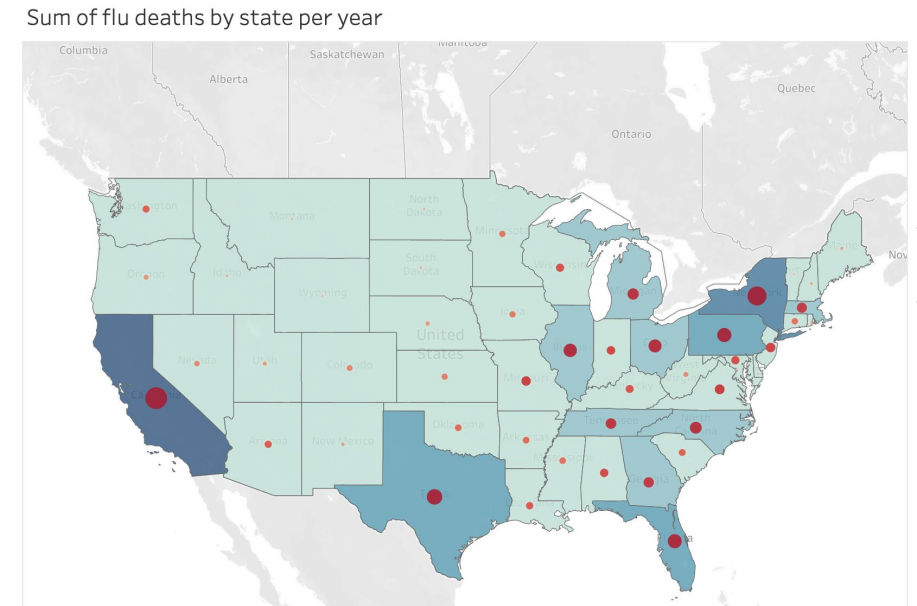
- Excel
- Translating business requirements
- Data cleaning
- Data integration
- Data transformation
- Statistical hypothesis testing
- Visual analysis
- Forecasting
- Storytelling in Tableau
- Presenting results to an audience

Objectives, data, and general overview

- **Objectives**:
  - Decide optimal staffing levels for healthcare providers in the US in advance of flu season
  - Discover who vulnerable populations are and where they are located
  - Determine when flu season happens
- **Data**:
  - US government census data
  - CDC flu death data
- **General overview**:
  - Certain populations across the US are more susceptible to adverse health outcomes when they contract the flu. To most effectively allocate additional healthcare personnel, a staffing agency wants to know who is most vulnerable, where these populations are located, and when flu prevalence is highest each year.

Findings


Correlation between age and deaths from influenza in population 65+ by state
Population aged 65+
4M
2M
0M
0K
1K
2K
3K
4K
5K
6K
Sum of deaths in population 65+


Sum of flu deaths by state per year
Columbia
Alberta
Saskatchewan
Ontario
Quebec
Montana
North Dakota
South Dakota
Wyoming
United States
New Mexico

Recommendations

- Since states with higher populations also have larger populations aged 65 and over — who are more vulnerable to the flu — staffing plans should take this into consideration
- Seasonality of the flu should also be factored in when creating staffing plans for allocating healthcare providers

Take a look at the Tableau storyboard for more in-depth information. You can also watch my presentation about the project on YouTube.

Answering business questions for an online video rental company

Rockbuster

- Relational databases
- SQL
- Database querying
- Filtering
- Cleaning and summarizing
- Joining tables
- Subqueries
- Common table expressions

Objectives, data, and general overview

- **Objectives**:
  - Help inform Rockbuster Stealth's BI department with the launch strategy for the new online video service
- **Data**:
  - Data set that contains information about Rockbuster's film inventory, customers, and payments, among other things.
- **General overview**:
  - Rockbuster Stealth LLC is a movie rental company that used to have stores around the world. Facing stiff competition from streaming services such as Netflix and Amazon Prime, the Rockbuster Stealth management team is planning to use its existing movie licenses to launch an online video rental service in order to stay competitive.

Findings


Rockbuster Customers Around the World by Revenue
120
560
3,685
2,985
224
632
98
369
407
2,919
206
303
273
1,299
140
97
334
284
514
205
74
274
349
659
115
123
203
186
169
122
315
93
1,069
162
2,766
5,251
474
6,035
353
105
2,220
104
330
1,353
3,123
144
94
65
86
© 2024 Mapbox © OpenStreetMap

Recommendations

- Tailor strategy to local preferences in terms of genre and rating
- Focus on countries with highest customer numbers: India, China, United States, Japan, Mexico, Brazil. Russia, Philippines, Turkey, and Indonesia
- Offer promotions to high lifetime value customers
- Determine appropriate entry price level based on average movie rental duration and average rental rates

Marketing strategy for an online grocery store

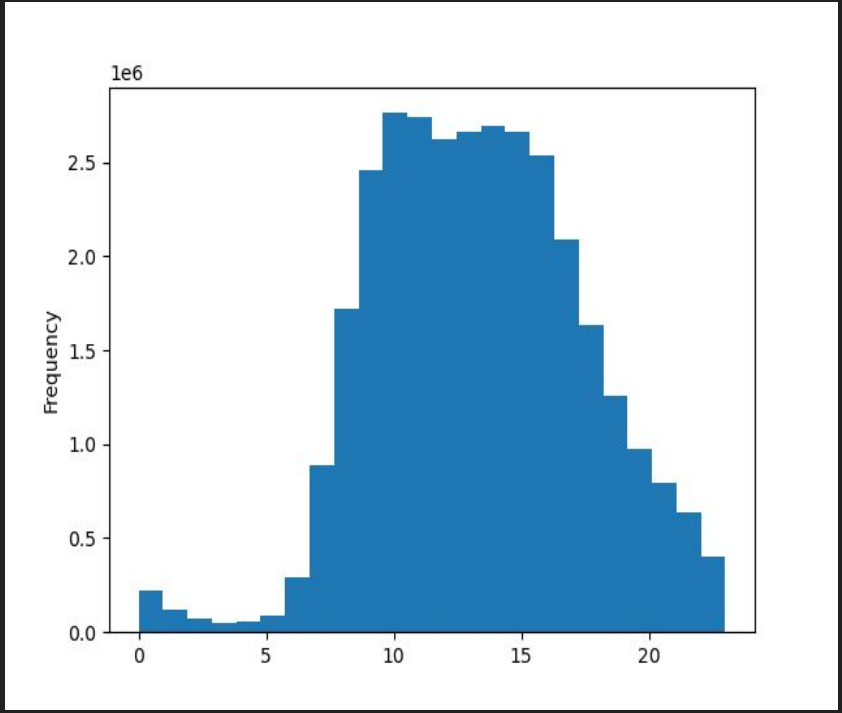
Instacart

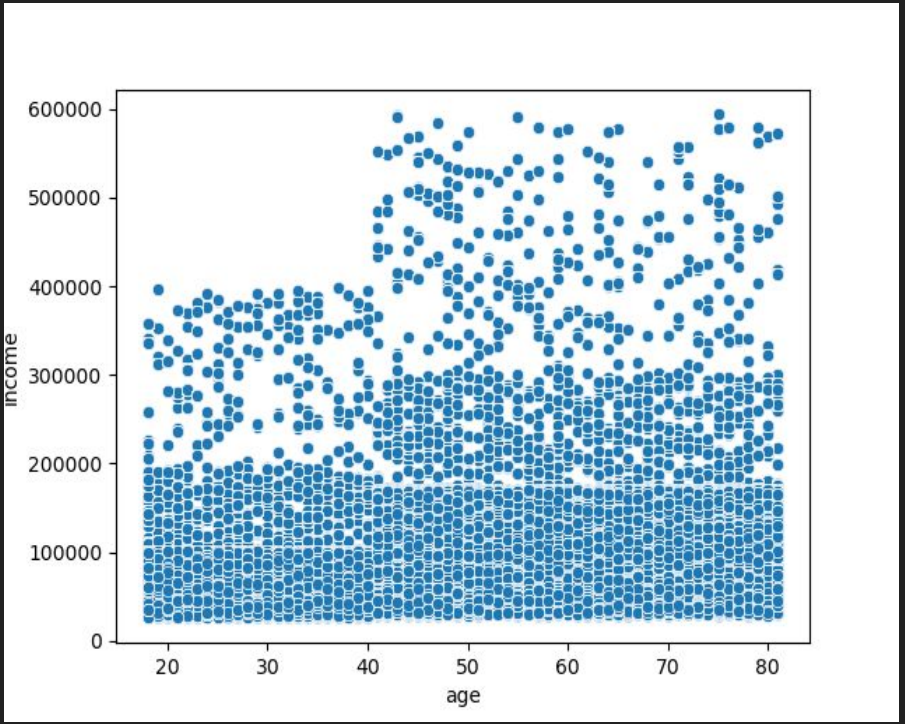
- Python
- Data wrangling
- Data merging
- Deriving variables
- Grouping data
- Aggregating data
- Reporting in Excel
- Population flows

Objectives, data, and general overview

- **Objectives**:
  - Analyze data to create data-driven marketing strategies
  - Discover new insights about ordering habits and create customer profiles
- **Data**:
  - Instacart product and orders datasets
  - Customer and price datasets
- **General overview**:
  - Instacart has a wealth of knowledge about their customers and their orders and wants to use this information to inform their marketing decisions.
  - Drilling down on customers and their orders based on region, family status, order time and day, repeating orders, and other factors can help fine-tune marketing strategy and increase sales.

Findings


1e6
2.5
2.0
1.5
1.0
0.5
0.0
Frequency
0
5
10
15
20


600000
500000
400000
300000
200000
100000
0
income
20
30
40
50
60
70
80
age

Recommendations

- Optimize advertising to target days (Tuesdays + Wednesdays) and times (23:00 - 06:00) when orders are lowest
- Create a simpler price segmentation for further analysis, given that most items purchased are under $15
- Offer incentives to returning customers (10+ orders) to turn them into loyal customers (over 40 orders)
- Highlight promotions to increase basket size

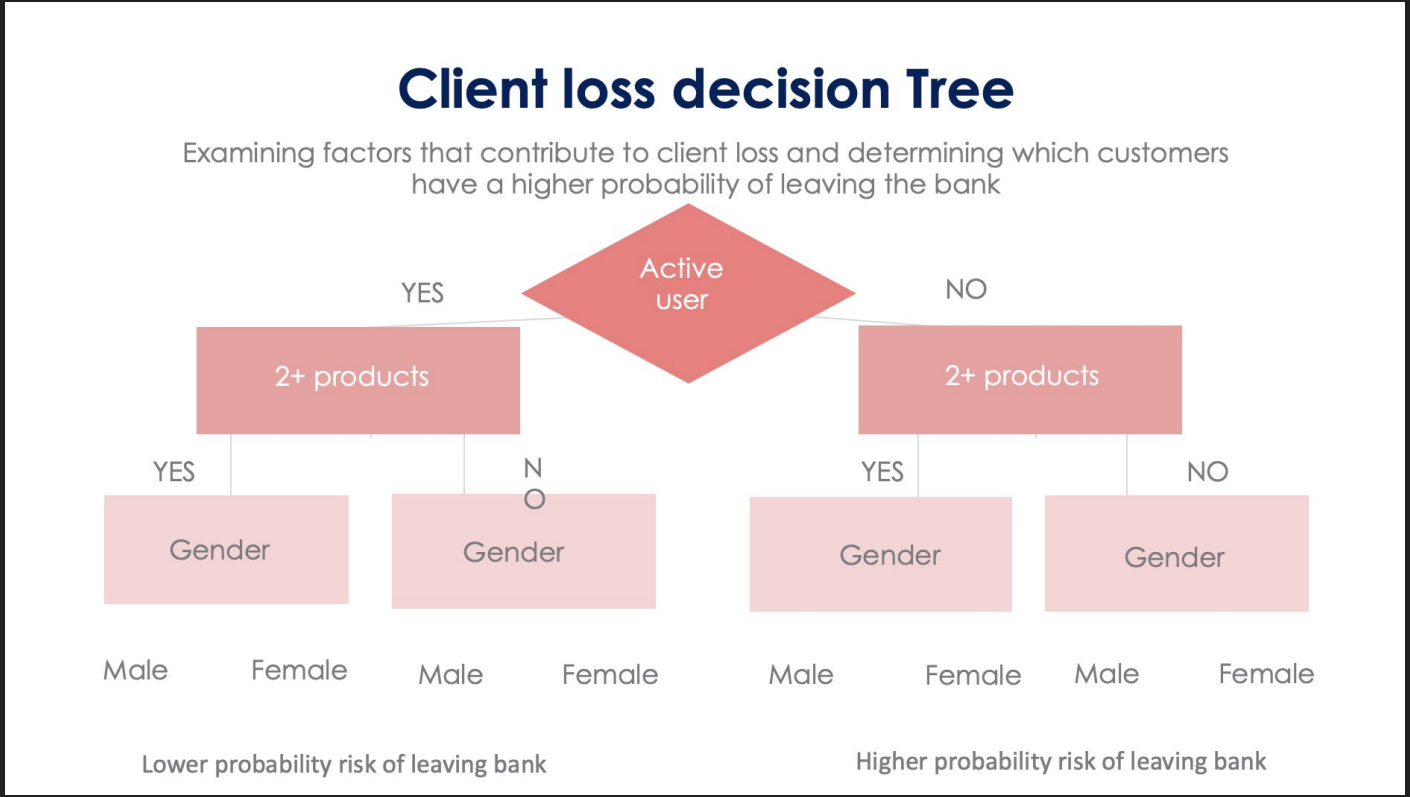
Anti-money laundering projects at a global bank

- Big data
- Data ethics
- Data mining
- Predictive analysis
- Time series analysis and forecasting
- Using GitHub

Objectives, data, and general overview

- **Objectives**:
  - Determine factors that can help identify customers who are most likely to leave bank, based on factors such as age, country, gender, amount of bank products they have, and whether they are active users or not, among others
- **Data**:
  - Data from Spain, France, and Germany, including customer age, country, gender, amount of products they have, activity status, credit scores, and more
- **General overview**:
  - The bank is noticing customers leaving the bank and wants to halt attrition. To do so, they want to discover any commonalities between customers who are choosing to stop banking with them.

Findings


Client loss decision Tree
Examining factors that contribute to client loss and determining which customers have a higher probability of leaving the bank
Active user
YES
NO
2+ products
2+ products
YES
NO
YES
NO
Gender
Gender
Gender
Gender
Male
Female
Male
Female
Male
Female
Male
Female
Lower probability risk of leaving bank
Higher probability risk of leaving bank

Recommendations

- Since active users are overall less likely to leave the bank, create ways to link customers to the bank more actively
- Women are more likely to leave the bank, so create programs to address them specifically
- Women in Germany are leaving the bank in higher percentages, so run a localized campaign to incentivize them to stay
- As customers with more products are less likely to leave the bank, create product bundles to enroll customers in more products

Analyzing NBA stats

Diving into NBA data from 1996-2022 to discover trends, correlations, and more

- Python
- Excel
- Data sourcing
- Data cleaning
- Data exploration
- Cluster analysis
- Correlation matrix
- Pair plots
- Regression analysis
- Cluster analysis
- Time series analysis

Objectives, data, and general overview

- **Objectives**:
  - Explore NBA data set to discover connections, correlations, and other interesting trends
  - Perform exploratory data analysis, clustering, and time series analysis
- **Data**:
  - NBA Players dataset, covering 1996 through 2022, including biographical data about players, as well as teams, the college they played for, and stats such as points, rebounds, assists, net rating, usage percentage, games played, and more
  - Additional data from Basketball Reference
- **General overview**:
  - NBA — and analytics — fans around the world can look more deeply at different developments in the games, the teams, and how the game is played
  - With an ever-increasing popularity in both the NBA in general and Fantasy Basketball leagues, more people than ever are keen to delve into statistical analysis around player performances on a more granular level

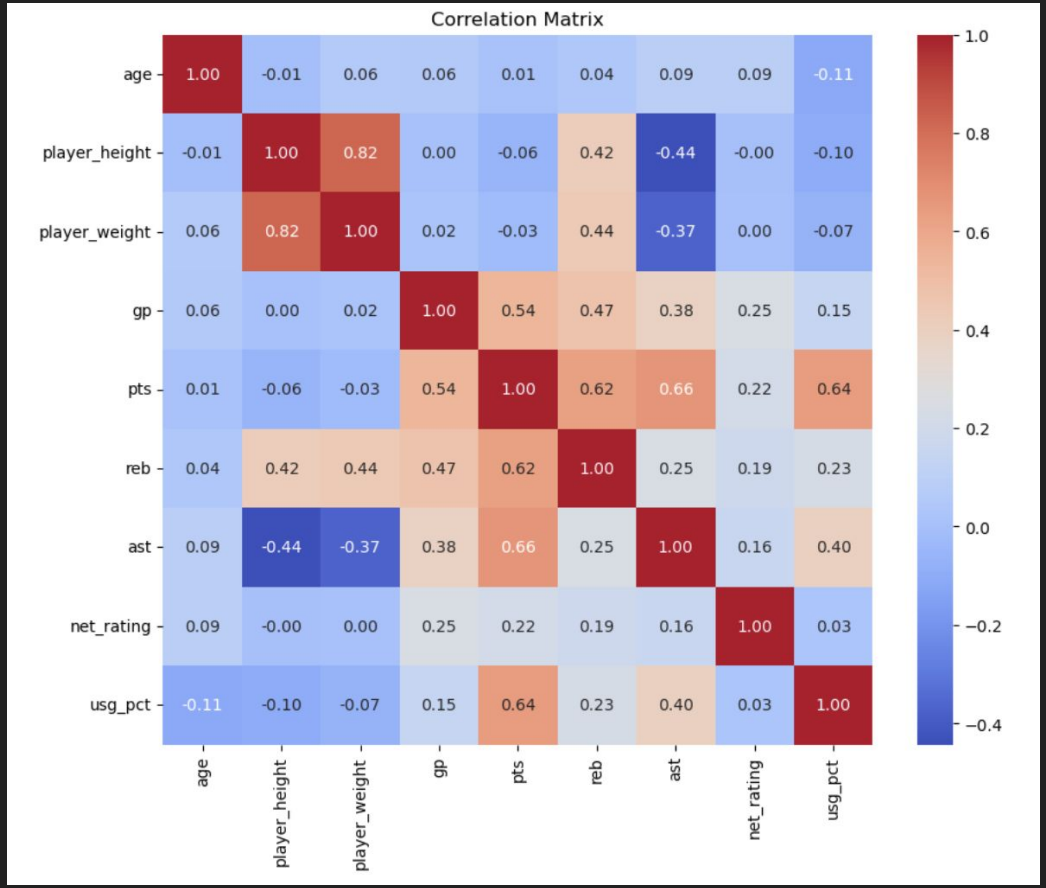
Findings

We can see positive relationships between:

- usage percentage and points (0.64) and usage percentage and assists (0.40)
- assists and points (0.66)
- rebounds and player height (0.42), rebounds and player weight (0.44), rebounds and games played (0.47), and rebounds and points (0.62)
- points and games played (0.54)
- player weight and player height (0.82)

There are negative correlations between assists and player height (-0.44) and assists and player weight (-0.37)

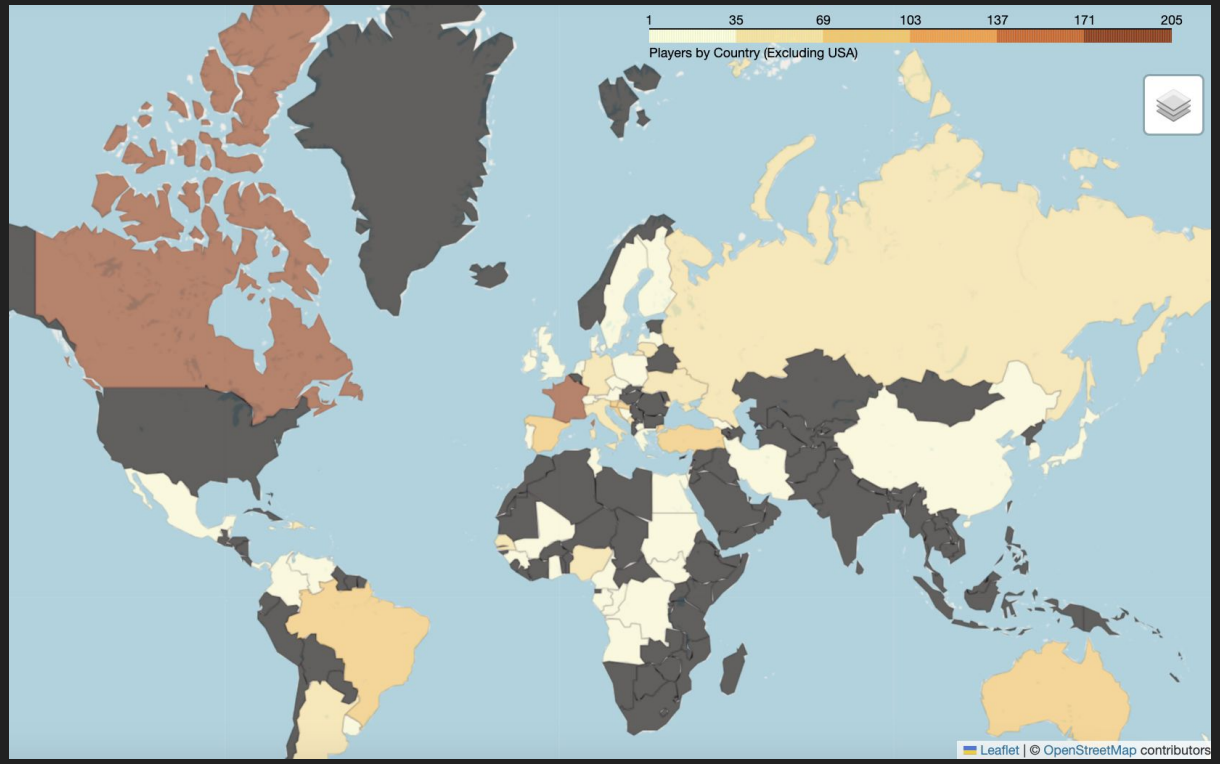
Unexpectedly, player age is almost in no way correlated with any of the other variables.



Findings

Geographical origin is predominantly from the USA, but over the season from 1996 to 2022, players from many other countries were also represented in the NBA. The countries outside the USA producing the most talent are:

- Canada
- France
- Australia
- Spain
- Brazil
- Turkey
- Slovenia
- Croatia
- Serbia
- Argentina



Recommendations

- Further analysis of the data set to expand on findings of the analysis carried out so far — especially around variables that haven't been examined together
- Continue cluster analysis with more clusters to discover unknown groupings within the players
- Expand the analysis to include data from earlier years to expand understanding underlying trends and developments
- For NBA Coaches and/or those involved in Fantasy Basketball: selecting without looking at a player's age, only their stats

Check out my Tableau dashboard for interactive charts and more visualizations. You can take a look at all my findings in my GitHub repository.