

6210 Visualization Group Project Final Report

Group 3 - Aarya Paudyal, Atithi Prasai, Divi Joshi, Siddharth Lakhani, Weiheng Ze, Yash Luharuka,
Zhengkai Zhou



Introduction:

Football/soccer is a global phenomenon with millions of fans around the world. It is one of the most popular sports and has a rich history of memorable moments and important events. The objective of the game is to score more goals than the opposing team by getting the ball into the opponent's net. Players use any part of their body except their hands and arms to move the ball, and they typically pass the ball to their teammates using their feet, head, or chest. In soccer, disciplinary actions can include receiving a red card for serious offenses such as violent conduct or deliberate handballs, or a yellow card for minor offenses such as unsporting behavior or time-wasting. A penalty kick is awarded to the attacking team if a defending player commits a foul within their own penalty area, while a penalty corner is awarded if a defending player commits a foul within their own circle. Lastly, a player is deemed offside if they are closer to the opposing team's goal than both the ball and the second-last defender when the ball is played to them, resulting in a free kick for the opposing team. These penalties and disciplinary actions are enforced to maintain fair play and sportsmanship in the game.

The FIFA World Cup is an international football/soccer tournament and one of the biggest sporting events in the world. Members of the Federation Internationale de Football Association [FIFA] compete every four years for the title of World Cup Champion. The FIFA World Cup is widely considered the most-watched sports event on the planet thanks to record-breaking viewership numbers including more than 3.5 billion viewers—roughly half the world's population—during the 2018 finals tournament. France emerged as the final victor of the 2018 FIFA World Cup, which took place in Russia. Qatar hosted the 2022 FIFA World Cup from November 21 to December 17, 2022, marking the first time the event has been hosted by a Middle Eastern country. The 2026 tournament will be jointly hosted by Mexico, Canada, and the United States, making Mexico the first three-time host.

Problem Statement:

Improve the fan experience and increase revenue at the FIFA World Cup by using visual dashboards to monitor food preferences and track metrics such as wait times, menu items, and sales data. Implement predictive analytics and artificial intelligence to improve inventory management and delivery times, reducing food wastage. This will enhance the overall fan experience, increase satisfaction, and boost revenue for the sports industry.

Assumptions:

1. A FIFA world cup match is taking place between Germany and Brazil.
2. The match commences at 9:00 pm and concludes at 11:38 pm.
3. 400 spectators are present at the stadium to watch the match, which is equivalent to approximately 40,000 spectators in the actual scenario.
4. Each spectator receives a ticket that contains information about their gender, date of birth, seat location, and the team they are supporting.
5. The stadium is segregated into four wings, namely East, West, North, and South.
6. Each ticket comprises a QR code (see *Appendix C*) that redirects the holder to the food ordering app when scanned. The menu includes 36 items, including snacks, appetizers, beverages, and desserts, allowing them to avoid long queues at stalls and not miss out on the match.
7. Customers can specify their dietary preferences, such as Vegan, Vegetarian, Non-Veg, and Gluten-Free, while placing their food orders.
8. The match experiences various events like goals, corner kicks, yellow cards, red cards, and penalties, which are being monitored.
9. The time of placing an order has a maximum deviation of 3 minutes from the time of the occurrence of an event.

Product Details:

Our product consists of 2 important parts:

1. The visualization dashboard that is to be used by vendors running restaurants in the stadium.
2. The app that the customers will be using to order their food.

Dashboard:

1. It consists of data on the food item helping the vendor keep an accurate inventory management hence reducing wastage of food.
2. The dietary preferences of the customers are already noted down when the customer orders through the app.
3. The data on playing teams and supported teams from the customers. (out of the ~40,000 supporters, how many are supporting the winning team and the losing team). This will help us throw more offers based on the ordering pattern and the match situations.
4. We have a visual on the important events of the match which helps get an idea on the customer ordering behavior after a goal is scored by their supporting team, or the half time, break, red card, etc.
5. The zone wise data on the dashboard is an important part considering that the vendor can throw multiple offers to customers as pop ups in the zones where there are less orders being placed. For example - our Machine learning system will throw Beer offers for zone B as they are placing less drinks orders and more food orders. So we can offer them combo offers.
6. The gender bifurcation data helps the system understand age and gender wise trend in the order trend and customer behavior. (see *Appendix D*)

Food Ordering App:

1. The app can be downloaded by the customer through a QR code provided on the ticket. This gives us accurate data on the seat location, gender, date of birth and food preferences (veg, Vegan, Non-veg, allergies, etc).
2. As the customer can order only through the app from their seat they don't waste any time by waiting in long queues.
3. The app will also show betting options (subject to specific country's/ city's, betting rules) where a bet win will get the customer a free food / drink item or a discount. (*see Appendix E*)

Data Gathering and Generation Process:

We have used secondary data sources to collect information on:

1. Stadium size and seating capacity of the stadium. This information was used to determine the appropriate sample size of spectators/ audience that can attend the game and how they are seated across the stadium.
2. Items available in the menu and the prices at which they were sold at the game. This information was used to curate the menu table and determine the right prices for each item.
3. Past FIFA world cup schedules and the locations where they were held. This information was used to decide the timeline of the match.
4. List of teams that have participated in the past World Cup matches. This information was used to pick up two participating teams.
5. Popular cuisines in host and the countries of the participating teams. This information was used to design the food menu and pricing strategies to cater to the tastes and preferences of the audience.

We utilized ChatGPT to create R codes to generate tables based on the assumptions and secondary information such that they resemble the real world datasets. (*see Appendix A*)

Units of Analysis:

The unit of analysis in this scenario is the individual spectator who attends the FIFA world cup match and receives a ticket with information about their gender, date of birth, seat location, and team they are supporting. The individual spectator is the primary focus of the study, and data will be collected and analyzed based on their behavior and preferences. (*see Appendix B*)

Metrics:

1. Audience Metrics:
 - a. Total number of spectators present at the stadium
 - b. Gender distribution of the spectators
 - c. Age distribution of the spectators
 - d. Distribution of spectators based on the team they are supporting
 - e. Distribution of spectators based on their seat location
2. Menu Metrics:
 - a. Total number of orders placed for each menu item
 - b. Dietary preference distribution of the customers
3. Order Metrics:
 - a. Total number of orders placed during the match

- b. Distribution of orders based on the time of the match and the occurrence of events (goals, corner kicks, yellow cards, red cards, and penalties)
 - c. Revenue generated through sales
4. Event Metrics:
- a. Number of goals scored by each team
 - b. Number of corner kicks taken by each team
 - c. Number of yellow cards and red cards issued to players
 - d. Number of penalties awarded and converted by each team

These metrics can help evaluate the performance of various aspects of the match, including audience engagement, food ordering and delivery process, and overall match experience. The data collected from these metrics can be used to identify areas of improvement and optimize the spectator experience for future matches.

Wrangling using R:

We have performed the below wrangling tasks using R:

1. We loaded four CSV files, which are stored in R data frames named order, menu, audience, and events, respectively, using the `read.csv()` function.
2. The `tidyR`, `dplyr`, and `lubridate` libraries were loaded using the `library()` function.
3. Using R code we checked for missing values and corrected data form for each data frame using the `str()` function to view the structure of each data frame and the `sum(is.na())` function to calculate the number of missing values.
4. We performed Data cleaning and preparation for the data on order, menu, audience, and events data frames using various functions such as `mdy()` to convert date formats, `as.numeric()` and `gsub()` to remove the dollar sign from menu prices, and `select()` to remove unnecessary columns.
5. We have used the `unite()` function to unite the `Order.date` and `Order.time` columns in the order data frame into a new column named "`Order_date_time`", which is then saved as a new CSV file named "`neworder.csv`". (attached to the zip file shared)
6. We have then used the `inner_join()` function to join the audience and order data frames based on the `Aud_ID` column.
7. We created a new column named "Age" and calculated the age for each audience using the `age_calc()` function from the `lubridate` package. We have used the `round()` function to round the age values to the nearest integer.
8. The combined data frame with the new "Age" column is saved as a new CSV file named "`Age.csv`".

Impact of our Dashboard on Concession Wait Times:

After using the product there will be a drastic reduction in Concession Wait Times:

1. As per a report titled "Stadium of the Future" by Oracle, sports venues can increase revenue by reducing wait times at concession stands. Long lines at concession stands can discourage fans from purchasing food and drinks, leading to missed sales opportunities for the venue.
2. By implementing technology such as mobile ordering and payment, as well as self-service kiosks and intelligent queuing, venues can reduce wait times and improve the overall fan experience, leading to increased revenue. The report highlights that venues can also utilise data analytics to optimise staffing and inventory management, further improving efficiency and revenue generation. (see *Appendix F*)

Future Scope and Limitations:

The sports arena industry has increasingly used food to enhance the fan experience and drive engagement and revenue. However, despite the improvements in food options, the process of gathering data and understanding the specific food preferences of individual fans during game-related events has proven to be a challenge for a long time.

Our product aims to solve this challenge by providing a data-based dashboard that can be used to connect the nuances of the specific sporting game to food preferences at those moments in time, helping vendors to better understand their customers, meet the demand and increase revenue.

To achieve this, we will provide a comprehensive dashboard that focuses on key metrics to optimise vendor offerings, prices, and marketing strategies, as well as features that engage the audience and encourage them to order more.

By using our platform, vendors can simplify their operations, boost sales, and improve the overall customer experience. We provide valuable information such as popular local cuisines, audience dietary preferences, frequency of orders during the entire football match, and which areas of the stadium are likely to order more, helping vendors keep up with demand and stay prepared. Overall, our product provides a solution that helps vendors better understand their customers and maximise their revenue potential.

However, we do feel that the following points serve as a limitation to the project:

1. The cost of sourcing authentic ingredients and partnering with local chefs may be high, which could limit the scope of the project.
2. The food safety regulations and requirements may vary across different countries and venues (3 different countries will be hosting the World cup 2026), which could pose a challenge to ensure that the food offerings meet the necessary standards.
3. Coordinating with local chefs and ensuring that they have the necessary equipment and resources to prepare and serve their cuisine at the venues could be a logistical challenge.

References:

- <https://www.statista.com/statistics/1298103/fifa-world-cup-qatar-stadium-capacities/>
- <https://www.statista.com/statistics/873177/fifa-world-cup-interest/?locale=en>
- <https://www.statista.com/topics/9211/2022-fifa-world-cup/#topicOverview>
- <https://www.sportco.io/article/greatest-stadiums-in-fifa-world-cups-history-635097>
- <https://web.archive.org/web/20150607235353/http://www.fifa.com/fifa-tournaments/classic-stadiums/index.html>
- <https://www.stadiumguide.com/tournaments/fifa-world-cup-2022-stadiums-qatar/>
- <https://www.foxsports.com.au/football/world-cup/fifa-breaks-silence-on-world-cup-stadium-capacity-furore/news-story/6251d10d99602d1ab34940edb8120be4>
- <https://www.planetfootball.com/quick-reads/world-cup-2026-stadiums-host-canada-mexico-usa-los-angeles-new-york/>
- <https://lifestyle.livemint.com/food/discover/the-high-cost-of-drinking-and-dining-at-the-fifa-world-cup-2022-111668822114564.html>
- <https://www.dohaguides.com/stadium-food-menu-and-prices/>
- <https://abcnews.go.com/Lifestyle/2014-fifa-world-cup-concession-food-brazil/story?id=24058078>
- https://www1.villanova.edu/villanova/law/academics/sportslaw/commentary/msli_blog/journal_archives/2014/0728.html
- <https://www.dazn.com/en-US/news/soccer/fifa-world-cup-hosts-list-of-every-country-that-has-ever-staged-the-famous-tournament/1oj5lmkizmhej1h2vb30u3kx09>
- <http://www.endlesssummer.com/2010/06/07/the-world-cup-of-food/>
- <https://digitalhub.fifa.com/m/6584e4777be1c323/original/s42d4os0bo0pb3q1val-pdf.pdf>
<https://www.gillettestadium.com/concessions/>
- <https://www.mbsfb.com/menus/fresh-mex>
- <https://www.mbsfb.com/>

Appendix A: R codes for data generation

1. Audience table:

```
#-----
#Audience table

#Create a sequence of IDs
ID <- 1:400

#Generate random names
names <- randomNames(400,which.names = "first")

#Generate random dates of birth between 1950-2005
DOB <- as.Date(sample(as.Date('1950-01-01'), as.Date('2005-01-01'), by="day"), 400, replace=TRUE))

#Create a Gender column with Male, Female and Other genders
Gender <- sample(c("Male","Female","Other"),400,replace=TRUE)

#Create a Dietary preference column with Vegan, Vegetarian, Non-Veg and Gluten-Free
Dietary_preference <- sample(c("Vegan","Vegetarian","Non-Veg","Gluten-Free"),400,replace=TRUE)

#Create a Supporting Team column with FIFA World Cup teams
Supporting_Team <- sample(c("Brazil","Germany"),400,replace=TRUE)

#Create a Zone column with East wing, West wing, North wing and South wing
Zone <- sample(c("East wing","West wing","North wing","South wing"),400,replace=TRUE)

#Combine all the columns into a data frame
Audience <- data.frame(ID,names,DOB,Gender,Dietary_preference,Supporting_Team,Zone)

#-----
```

2. Event table:

```
#-----
#Events table

# Create a sequence of alphanumeric Event_ID
Event_ID <- paste0("E",seq(1:20))

# Create an Event column with events according to the given time intervals
Event <- c("Goal","Corner Kick","Yellow Card","Red Card","Penalty Kick","Halftime","Goal",
"Corner Kick","Yellow Card","Red Card","Penalty Kick","Break","Goal","Goal","Corner Kick","Break",
"Break","Break","Goal","NA")

# Create a Date column set to 2023-03-27
Date <- rep(as.Date("2023-03-27"),20)

# Create a Time column with events according to the given time intervals
Time <- c("21:03:23","21:13:12","21:21:45","21:30:59","21:41:32","21:45:00","22:01:23",
"22:11:01","22:19:47","22:26:23","22:37:56","22:45:00","23:02:30","23:05:23","23:12:59",
"23:15:00","23:16:39","23:19:59","23:21:23","23:35:00")

#Create Team column with team names
Team <- sample(c("Brazil","Germany"),20,replace=TRUE)

# Combine all the columns into a data frame
Events <- data.frame(Event_ID,Event,Date,Time,Team)

#-----
```

3. Menu table:

```
#-----
#Menu table

# create a vector of item names
item_names <- c("Feijoada", "Churrasco", "Schnitzel", "Bratwurst", "Hot Dog",
               "Hamburger", "Veggie Burger", "Veggie Wrap", "BBQ Chicken",
               "Chicken Wings", "Caesar Salad", "Greek Salad", "Nachos",
               "Popcorn", "French Fries", "Veggie Pizza", "Pepperoni Pizza",
               "Cheese Pizza", "Sausage Pizza", "Beer (small)", "Beer (medium)",
               "Beer (large)", "Soft drink (small)", "Soft drink (medium)",
               "Soft drink (large)", "Bottled Water", "Lemonade", "Ice Cream Sundae",
               "Strawberry Milkshake", "Chocolate Milkshake", "Mango Lassi",
               "Pretzels", "Cheeseburger", "Chicken sandwich", "Onion rings",
               "Falafel Plate")

# create a vector of prices
prices <- c(15.99, 18.99, 12.99, 9.99, 7.99, 8.99, 10.99, 9.99, 12.99,
           9.99, 7.99, 8.99, 6.99, 4.99, 3.99, 12.99, 14.99, 13.99,
           11.99, 5.99, 6.99, 7.99, 2.99, 3.99, 4.99, 1.99, 2.99, 6.99,
           7.99, 7.99, 4.99, 5.99, 4.99, 6.99, 8.99, 5.99)

# create a vector of food categories
food_categories <- c("non-veg", "non-veg", "non-veg", "non-veg", "non-veg",
                      "non-veg", "vegan", "vegan", "non-veg", "non-veg",
                      "vegan", "vegan", "non-veg", "vegan", "vegan", "vegan",
                      "non-veg", "non-veg", "non-veg", "non-veg", "non-veg",
                      "non-veg", "non-veg", "non-veg", "non-veg", "vegan",
                      "vegan", "vegan", "vegan", "vegan", "vegan", "non-veg",
                      "non-veg", "non-veg", "vegan")

# create the 'Menu' table
Menu <- data.frame(Menu_ID = 1:length(item_names),
                    Item_name = item_names,
                    Price = prices,
                    Food_Category = food_categories)
-----
```

4. Order table:

```
#-----
#Order table

# load necessary libraries
library(tidyverse)
library(lubridate)

# generate Order_ID column
Order_ID <- seq_len(800)

# generate Aud_ID column by sampling from Audience table's ID column
Aud_ID <- sample(Audience$ID, size = 800, replace = TRUE)

# generate Item_ID column by sampling from Menu table's Menu_ID column
Item_ID <- sample(Menu$Menu_ID, size = 800, replace = TRUE)

# generate Event_ID column by sampling from Events table's Event_ID column
Event_ID <- sample(Events$Event_ID, size = 800, replace = TRUE)

# generate Order date column
Order_date <- as.Date("2023-03-27")

# generate Order time column with a maximum deviation of 3 minutes from the event time
Order_time <- sapply(Event_ID, function(x) {
  event_time <- Events %>% filter(Event_ID == x) %>% select(Time) %>% unlist()
  event_time <- strptime(paste(Order_date, event_time), "%Y-%m-%d %H:%M:%S")
  event_time <- event_time + runif(1, -180, 180)
  format(event_time, "%H:%M:%S")
})

# generate Quantity column with random values between 1 and 5
Quantity <- sample(1:5, size = 800, replace = TRUE)

# combine all columns into a data frame
Order <- data.frame(Order_ID, Aud_ID, Item_ID, Event_ID, Order_date, Order_time, Quantity)
-----
```

Appendix B: Generated Tables

Audience

ID	names	DOB	Gender	Dietary_preference	Supporting_Team	Zone
1	Daysy	1999-09-25	Male	Gluten-Free	Germany	West wing
2	Gaitha	1964-04-15	Other	Vegan	Brazil	West wing
3	Gabriel	1950-05-16	Male	Vegan	Brazil	West wing
4	Sabrena	1950-11-11	Male	Vegetarian	Brazil	North wing
5	Silvia	1991-04-16	Male	Vegan	Brazil	West wing
6	Arvind	1956-05-21	Other	Non-Veg	Germany	South wing
7	Akane	1966-07-25	Female	Gluten-Free	Brazil	West wing
8	Sanjana	1980-07-06	Male	Vegetarian	Germany	North wing
9	Keyondra	1953-09-06	Other	Vegetarian	Germany	North wing
10	Shannon	1981-07-13	Male	Vegetarian	Germany	South wing

Event

Event_ID	Event	Date	Time	Team
E1	Goal	2023-03-27	21:03:23	Brazil
E2	Corner Kick	2023-03-27	21:13:12	Germany
E3	Yellow Card	2023-03-27	21:21:45	Germany
E4	Red Card	2023-03-27	21:30:59	Germany
E5	Penalty Kick	2023-03-27	21:41:32	Germany
E6	Halftime	2023-03-27	21:45:00	Germany
E7	Goal	2023-03-27	22:01:23	Germany
E8	Corner Kick	2023-03-27	22:11:01	Germany
E9	Yellow Card	2023-03-27	22:19:47	Germany
E10	Red Card	2023-03-27	22:26:23	Germany
E11	Penalty Kick	2023-03-27	22:37:56	Germany
E12	Break	2023-03-27	22:45:00	Germany
E13	Goal	2023-03-27	23:02:30	Brazil
E14	Goal	2023-03-27	23:05:23	Brazil
E15	Corner Kick	2023-03-27	23:12:59	Germany
E16	Break	2023-03-27	23:15:00	Brazil
E17	Break	2023-03-27	23:16:39	Brazil
E18	Break	2023-03-27	23:19:59	Germany
E19	Goal	2023-03-27	23:21:23	Brazil
E20	NA	2023-03-27	23:35:00	Germany

Menu

Menu_ID	Item_name	Price	Food_Category
1	Feijoada	15.99	non-veg
2	Churrasco	18.99	non-veg
3	Schnitzel	12.99	non-veg
4	Bratwurst	9.99	non-veg
5	Hot Dog	7.99	non-veg
6	Hamburger	8.99	non-veg
7	Veggie Burger	10.99	vegan
8	Veggie Wrap	9.99	vegan
9	BBQ Chicken	12.99	non-veg
10	Chicken Wings	9.99	non-veg
11	Caesar Salad	7.99	vegan
12	Greek Salad	8.99	vegan
13	Nachos	6.99	non-veg
14	Popcorn	4.99	vegan
15	French Fries	3.99	vegan
16	Veggie Pizza	12.99	vegan
17	Pepperoni Pizza	14.99	non-veg
18	Cheese Pizza	13.99	non-veg
19	Sausage Pizza	11.99	non-veg
20	Beer (small)	5.99	non-veg
21	Beer (medium)	6.99	non-veg
22	Beer (large)	7.99	non-veg
23	Soft drink (small)	2.99	non-veg
24	Soft drink (medium)	3.99	non-veg
25	Soft drink (large)	4.99	non-veg
26	Bottled Water	1.99	vegan
27	Lemonade	2.99	vegan
28	Ice Cream Sundae	6.99	vegan
29	Strawberry Milkshake	7.99	vegan
30	Chocolate Milkshake	7.99	vegan
31	Mango Lassi	4.99	vegan
32	Pretzels	5.99	non-veg
33	Cheeseburger	4.99	non-veg
34	Chicken sandwich	6.99	non-veg
35	Onion rings	8.99	vegan
36	Falafel Plate	5.99	vegan

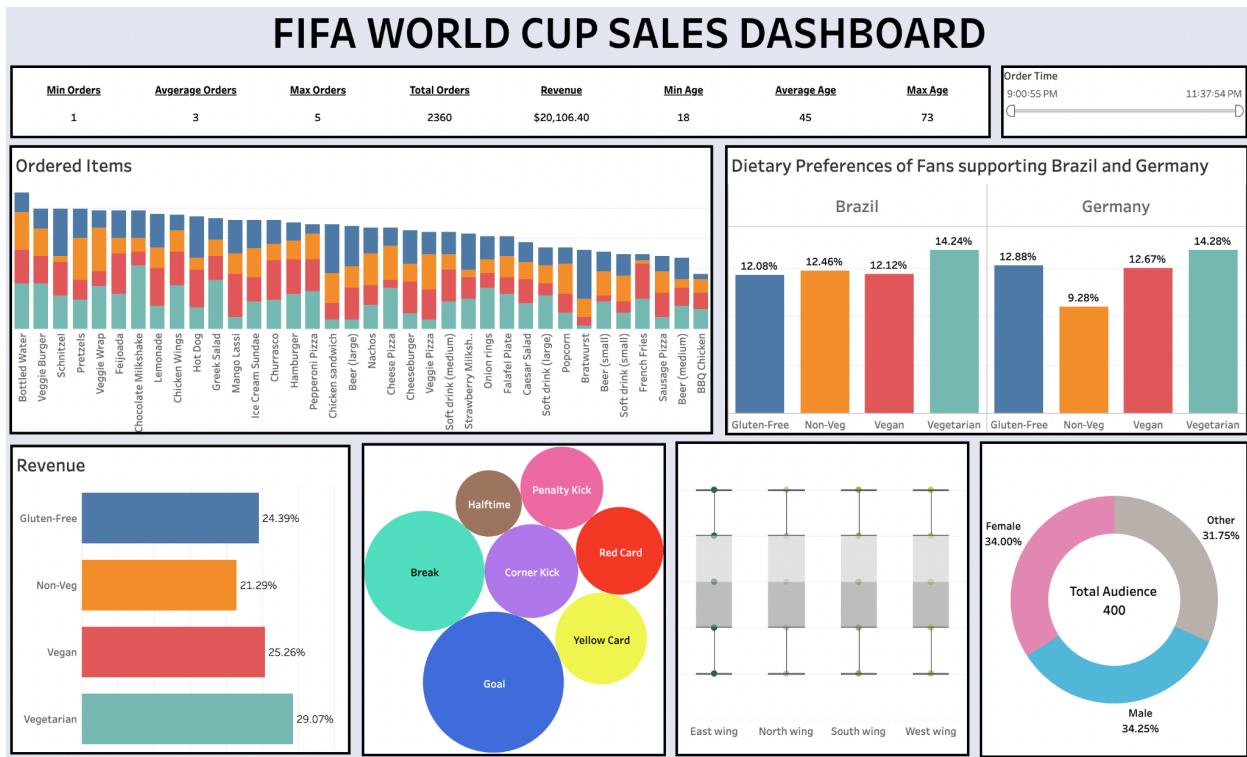
ordernew

Order_ID	Aud_ID	Item_ID	Event_ID	Order_date_time	Quantity
1	324	8	E3	2023-03-27 21:20:21	4
2	106	33	E7	2023-03-27 22:04:21	5
3	237	29	E8	2023-03-27 22:12:59	2
4	78	15	E5	2023-03-27 21:40:04	3
5	162	33	E17	2023-03-27 23:18:41	2
6	336	14	E10	2023-03-27 22:27:20	4
7	399	31	E10	2023-03-27 22:27:33	1
8	400	2	E15	2023-03-27 23:11:57	4
9	275	11	E16	2023-03-27 23:13:26	5
10	389	25	E16	2023-03-27 23:14:43	2

Appendix C: Sample Stadium ticket having QR code

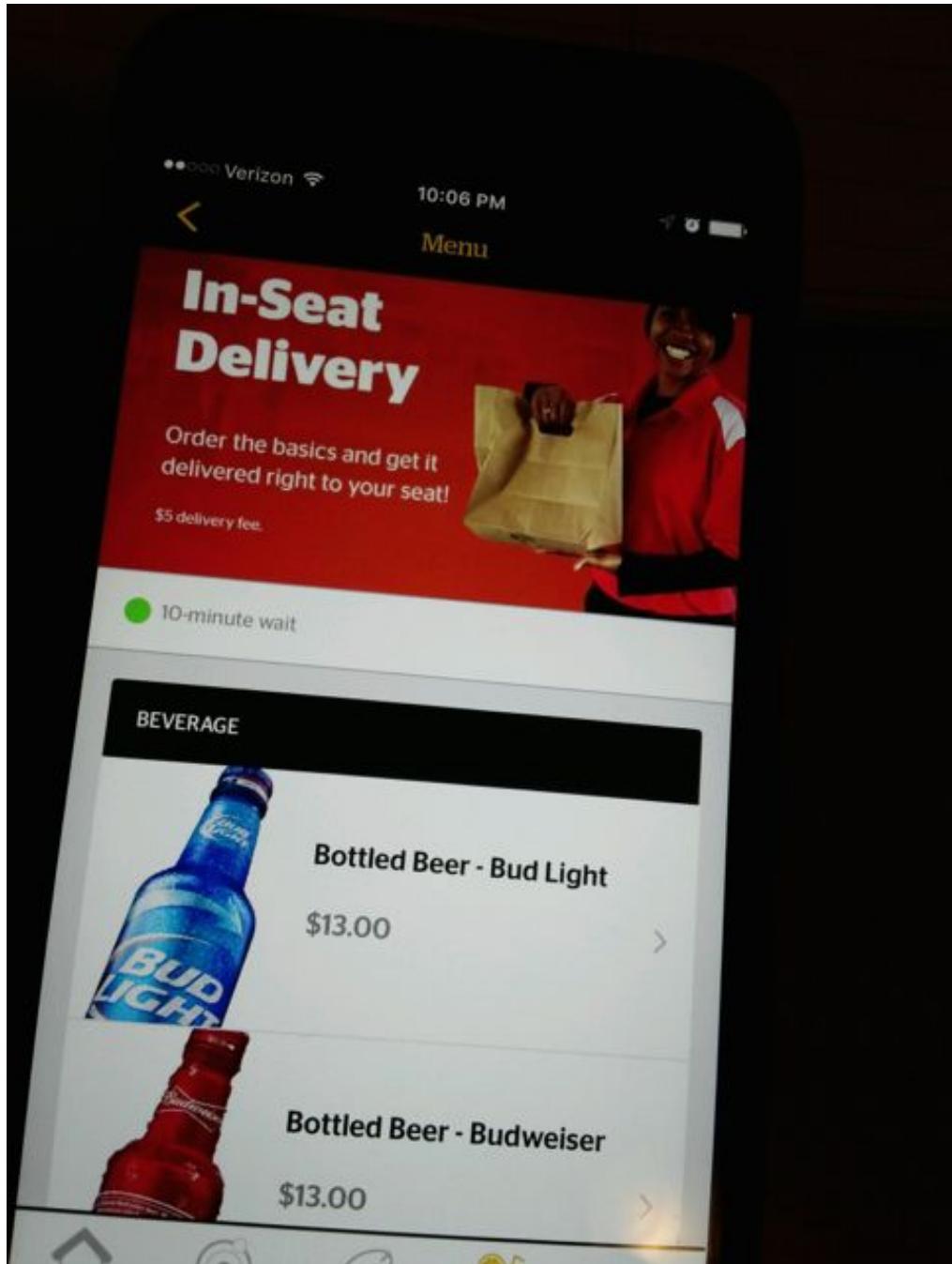


Appendix D: Tableau Dashboard



https://public.tableau.com/app/profile/divi.joshi/viz/Revised_16819069640000/Dashboard?publish=yes

Appendix E: Food Ordering App



Appendix F:

OPPORTUNITY TO INCREASE REVENUE BY REDUCING CONCESSIONS WAIT TIMES

	Average Spend at Concessions	Increase in Spend if Wait Time Cut in Half	Potential Increase in Fan Spend
Australia	\$32.41	\$13.59	41.93%
Brazil	R\$48.76	R\$ 24.13	49.49%
China	¥56.83	¥24.53	43.16%
France	€18.62	€9.38	50.38%
Germany	€20.42	€10.78	52.79%
Japan	¥2,195.04	¥670.04	30.53%
UK	£13.69	£6.10	44.56%
US	\$47.31	\$19.67	41.57%