Analyzing the WNBA

SQL Database Project

Rachel McDade

Background & Project Overview

Database to store and analyze data from the Women's National Basketball Association (WNBA)

Player, coach, and team statistics

Allows fans and viewers to dive into sports analytics

Why the WNBA?

I’m a fan!

Contribute to the field of women’s sports

Learn more about sports analytics

Project Objectives

Design WNBA database using PGAdmin and PostreSQL

Collect, clean, and input data

Data manipulation & querying

Python integration

Data visualization with Tableau

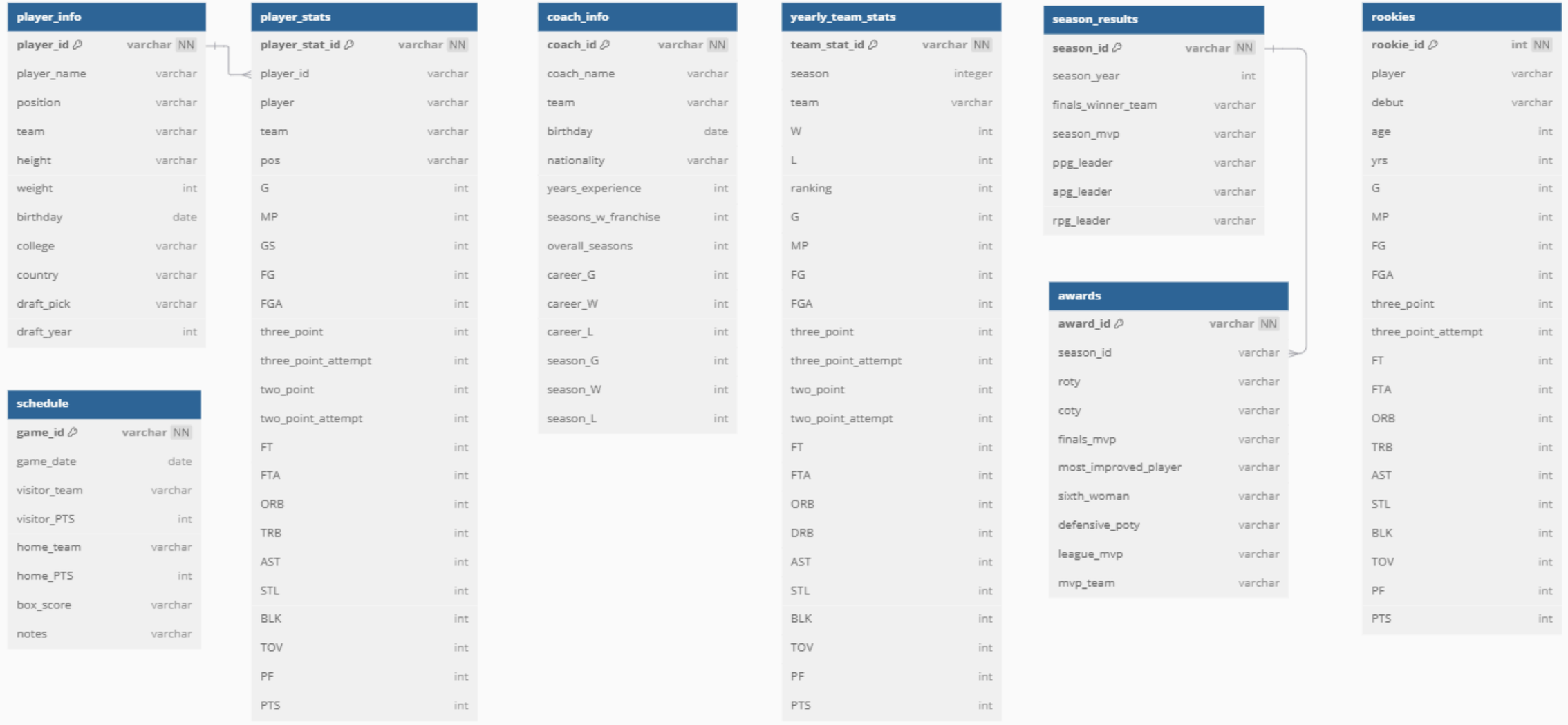
Project Walkthrough

Database Creation

pgAdmin

PostgreSQL

Relational Schema



Database Implementation & Data Insertion

DDL Statements

SQL INSERT Statements

Data from online sources

```
-------------------------------------------------------
/*                          Database Creation
-------------------------------------------------------
-- Create wnba database
CREATE DATABASE wnba;

-------------------------------------------------------
/*                          Player Info Table
-------------------------------------------------------
-- Create table player_info
CREATE TABLE player_info(
    player_id varchar(10) PRIMARY KEY,
    player_name varchar(100) NOT NULL,
    pos varchar(50),
    team varchar(100),
    height varchar(10),
    weight integer,
    birthday date,
    college varchar(100),
    country varchar(100),
    draft_pick varchar(50),
    draft_year integer
);
```

```
--Insert data from csv into player_info table
COPY player_info
FROM 'C:\Users\Public\player_info.csv'
WITH (FORMAT CSV, HEADER);


--Insert data from csv into player_stats table
COPY player_stats
FROM 'C:\Users\Public\player_stats.csv'
WITH (FORMAT CSV, HEADER);


--Insert data from csv into coach_info table
COPY coach_info
FROM 'C:\Users\Public\coach_info.csv'
WITH (FORMAT CSV, HEADER);


--Insert data from csv into yearly_team_stats table
COPY yearly_team_stats
FROM 'C:\Users\Public\yearly_team_stats.csv'
WITH (FORMAT CSV, HEADER);
```

Data Manipulation

Fill null values

Delete extraneous columns

Update abbreviations

Create new columns from existing data

```sql
-- Fill missing values in draft year with 0 to represent no draft year
--Identification
SELECT * FROM player_info
WHERE draft_year IS NULL;

--Update
UPDATE player_info
SET draft_year = 0
WHERE draft_year IS NULL;

-- Fill missing values in college with 'Unknown'
--Identification
SELECT * FROM player_info
WHERE college IS NULL;
--Update
UPDATE player_info
SET college = 'Unknown'
WHERE college IS NULL;
```

Query Examples

Demonstrated value of the database through query examples:

International mix of players

Information and stats on all #1 draft picks

Winning % of experienced vs non-experienced coaches

```
-------------------------------------------------------------------------------
/*                                Query 1                                    */
-------------------------------------------------------------------------------
--Count how many WNBA players are from each represented country
--Display result in alphabetical order by country
SELECT country, COUNT(country)
FROM player_info
GROUP BY country
ORDER BY country;

-------------------------------------------------------------------------------
/*                                Query 2                                    */
-------------------------------------------------------------------------------
--View information about #1 draft picks who are current players
--use an inner join to pull shooting percentage of each #1 draft pick
--order by draft year
SELECT
    pi.player_name AS player,
    pi.pos AS position,
    pi.team AS current_team, --specify current team vs team drafted to
    pi.draft_pick,
    pi.draft_year,
    ps.fg_pct
FROM player_info pi
INNER JOIN player_stats ps
ON pi.player_id = ps.player_id
```

Python Integration

Connected SQL database to Python with psycopg2 and pandas

Performed data analysis

- Descriptive statistics
- Data exploration

Data visualization with matplotlib and seaborn

- Distribution of data
- Relationships

Connecting the database to Python

```
# import libraries
import psycopg2
import pandas as pd
```

```
# Connect to wnba database
conn = psycopg2.connect("dbname=wnba user=postgres password=pgtour")
```

```
# Open a cursor to perform database operations
cur = conn.cursor()
```

```
# run a sql query using pandas for each table in the wnba database
# select all data from each table

# player_info table
player_info_query = pd.read_sql_query('''
                                      select * from player_info
                                      '''
                                      ,conn) # add conn variable that stores database connection info

# create a dataframe from the sql query
player_info = pd.DataFrame(player_info_query)

# view the df
player_info
```

Python Integration

Connected SQL database to Python with psycopg2 and pandas

Performed data analysis

- Descriptive statistics
- Data exploration

Data visualization with matplotlib and seaborn

- Distribution of data
- Relationships

Descriptive Statistics

```
# view descriptive stats for player_info
# table includes almost all categorical variables

player_info.describe()
player_info.describe(include='object')
```

	player_id	player_name	pos	team	height	birthday	college	country	draft_pick
count	142	142	142	142	142	142	142	142	142
unique	142	142	7	12	16	140	56	17	28
top	P3	Tina Charles	Guard	Washington Mystics	61	2001-02-11	Connecticut	USA	Rnd 1 Pick 1
freq	1	1	64	13	17	2	17	111	13

Observations:

Player Info

The median draft year is 2019, meaning half of the current players in the WNBA have been drafted since 2019. As this is only five years ago, we can see that many of the players are relatively inexperienced. Further the lower quartile draft year is 2015 (9 years ago), meaning 75% of players have been drafted since 2015. A basketball player interested in the WNBA could reasonably expect to have a short playing career.

Python Integration

Connected SQL database to Python with psycopg2 and pandas

Performed data analysis

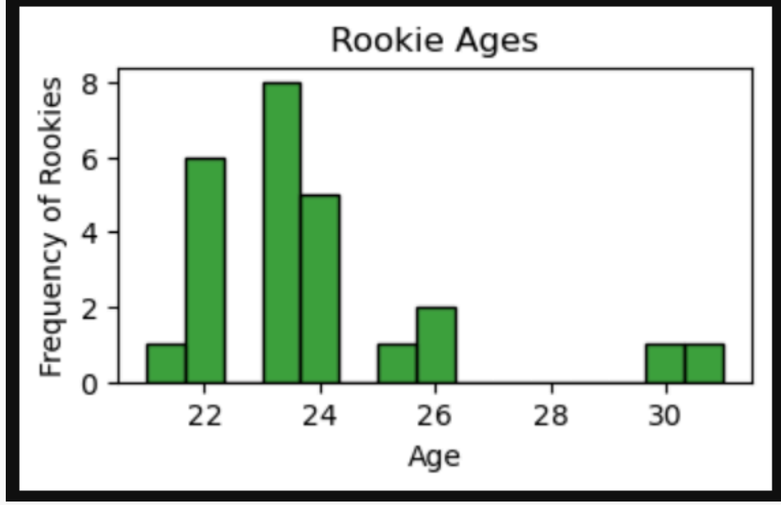
- Descriptive statistics
- Data exploration

Data visualization with matplotlib and seaborn

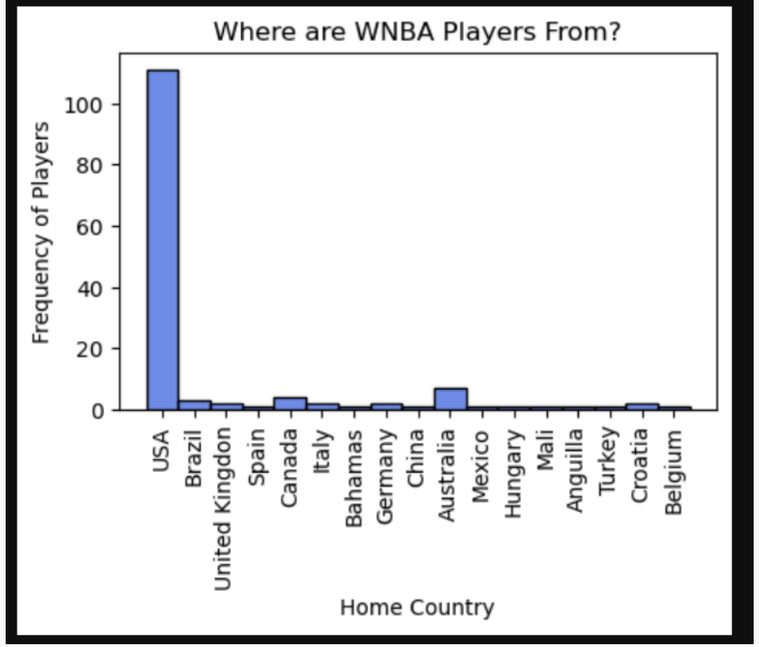
- Distribution of data
- Relationships

Data viz with matplotlib and seaborn

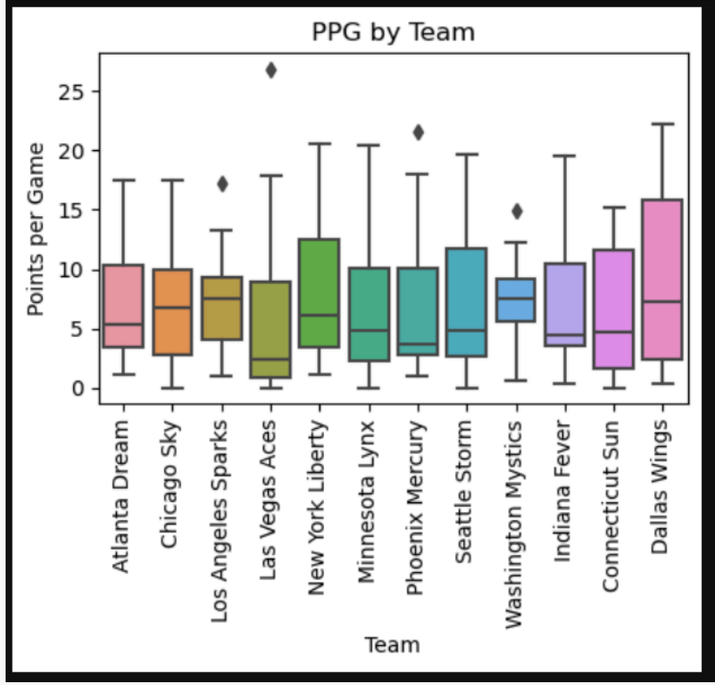
```python
# create histogram of rookie ages
plt.figure(figsize=(4, 2))
sns.histplot(rookies['age'], bins=15, color='green')
plt.title('Rookie Ages')
plt.xlabel('Age')
plt.ylabel('Frequency of Rookies')
plt.show()
```




```python
# create histogram player countries
plt.figure(figsize=(5, 3))
sns.histplot(player_info['country'], bins=20, color='royalblue')
plt.title('Where are WNBA Players From?')
plt.xlabel('Home Country')
plt.ylabel('Frequency of Players')
plt.xticks(rotation=90)
plt.show()
```



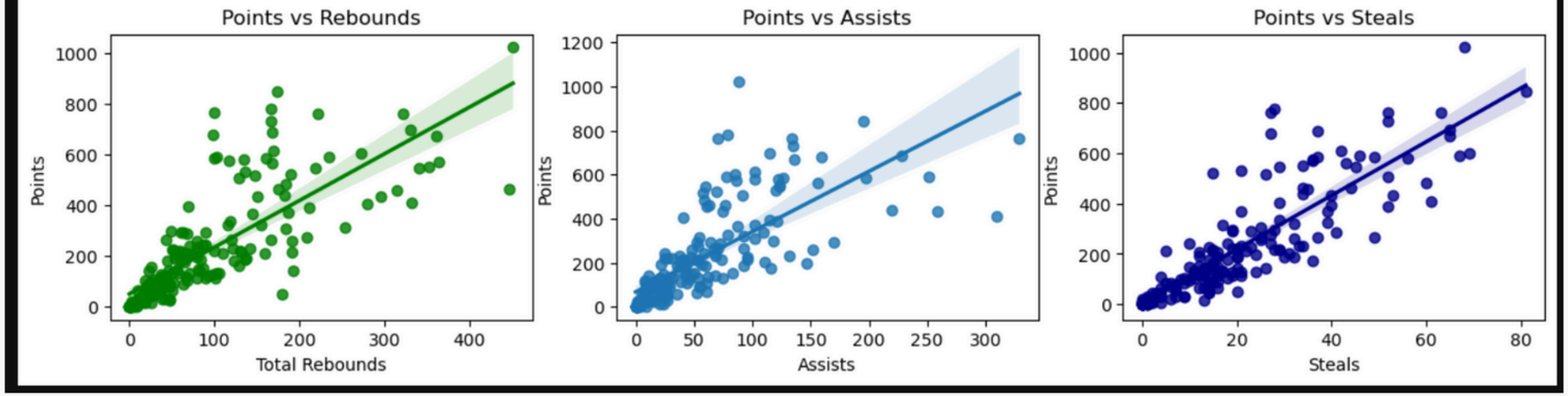

```python
plt.figure(figsize=(5, 3))
sns.boxplot(x='team', y='ppg', data=player_stats_pergame_filt_teams)
plt.title('PPG by Team')
plt.xlabel('Team')
plt.ylabel('Points per Game')
plt.xticks(rotation=90)  # Rotate x-axis labels for better readability
plt.show()
```



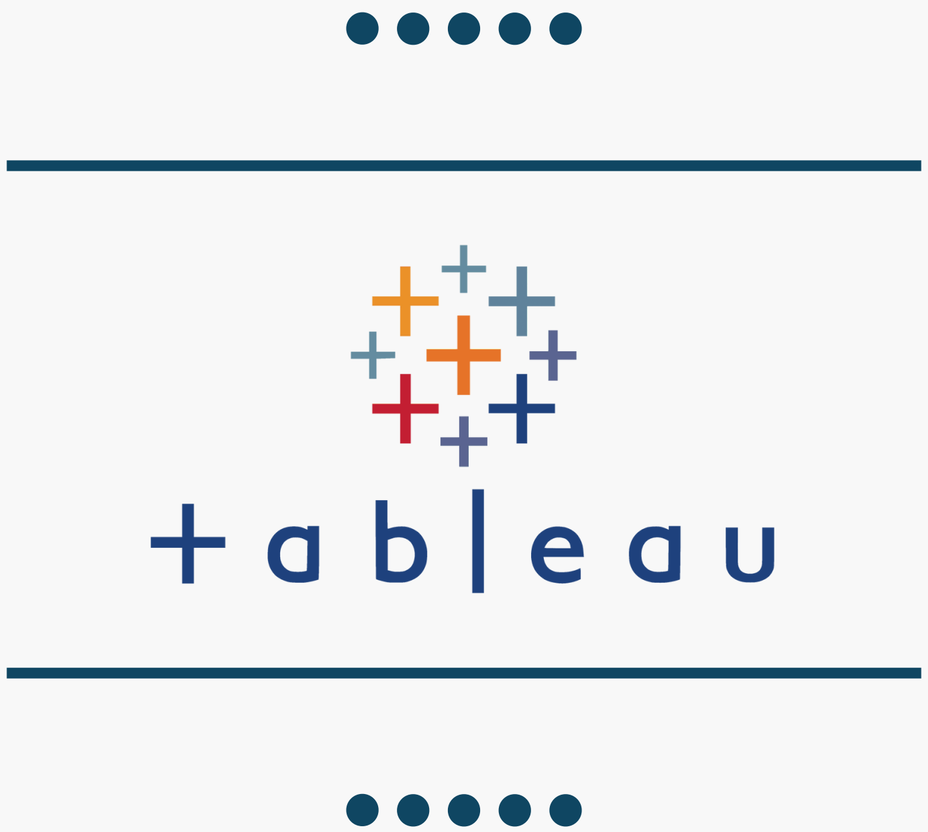
Data viz with matplotlib and seaborn

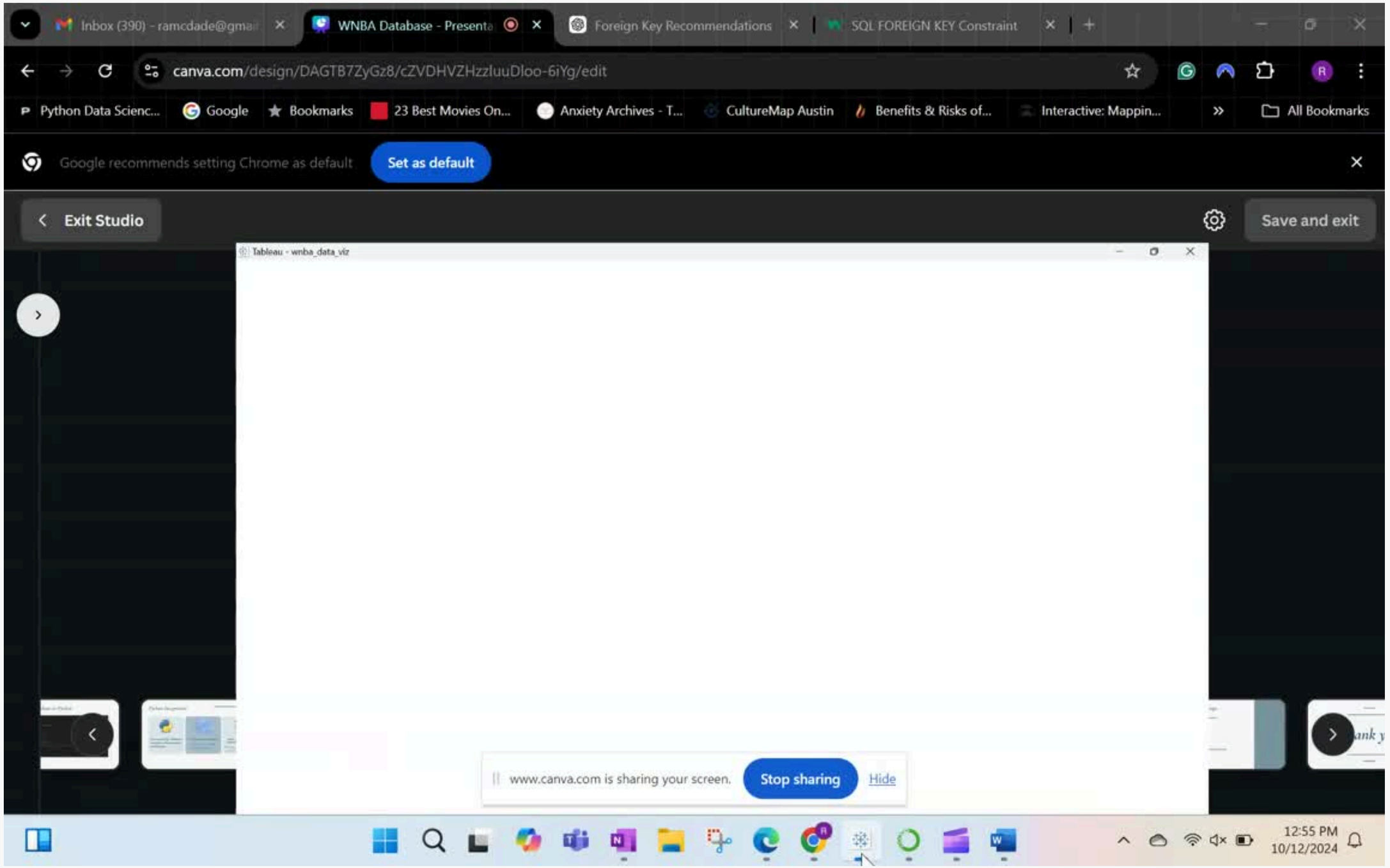
```
# steals scatterplot
plt.subplot(1, 3, 3)
sns.regplot(data=player_stats, x='stl', y='pts', color='darkblue')
plt.title('Points vs Steals')
plt.xlabel('Steals')
plt.ylabel('Points')
```

```
Text(0, 0.5, 'Points')
```



Data Visualization




Inbox (390) - ramcdade@gma
WNBA Database - Presenta
Foreign Key Recommendations
SQL FOREIGN KEY Constraint
canva.com/design/DAGTB7ZyGz8/cZVDHVZHzzluuDloo-6iYg/edit
Python Data Scienc...
Google
Bookmarks
23 Best Movies On...
Anxiety Archives - T...
CultureMap Austin
Benefits & Risks of...
Interactive: Mappin...
All Bookmarks
Google recommends setting Chrome as default
Set as default
Exit Studio
Save and exit
Tableau - wnba_data_viz
www.canva.com is sharing your screen.
Stop sharing
Hide
12:55 PM
10/12/2024

Challenges & Learnings

Allow more time for data collection

●●●●●

Narrow the focus

●●●●●

Try more alternative and/or new methods

●●●●●

Thank you