ADAN 8888 Project

September 23, 2024

[1]:
```
# Week Start: Ingestion of the Dataset
import pandas as pd

# File paths for each year
file_2019 = 'FFRank 2019.csv'
file_2020 = 'FFRank 2020.csv'
file_2021 = 'FFRank 2021.csv'
file_2022 = 'FFRank 2022.csv'
```

[2]:
```
# Read each CSV file into a pandas DataFrame
df_2019 = pd.read_csv(file_2019)
df_2020 = pd.read_csv(file_2020)
df_2021 = pd.read_csv(file_2021)
df_2022 = pd.read_csv(file_2022)
```

[3]:
```
# Display the first few rows of each dataset to ensure they loaded correctly
print("2019 Data Preview:\n", df_2019.head())
```

```
2019 Data Preview:
    Rank                 Player Team Position  Age  Games Played  \
0     1  Christian McCaffrey*+  CAR       RB   23            16
1     2        Lamar Jackson*+  BAL       QB   22            15
2     3         Derrick Henry*  TEN       RB   25            15
3     4            Aaron Jones  GNB       RB   25            16
4     5       Ezekiel Elliott*  DAL       RB   24            16

   Passing Completion  Passing Attempts  Passing Yards  Passing TDs  …  \
0                   0                 2              0            0  …
1                 265               401           3127           36  …
2                   0                 0              0            0  …
3                   0                 0              0            0  …
4                   0                 0              0            0  …

   Rushing TDs  Targets  Recepotions  Receiving Yards  Yards per Reception  \
0           15      142          116             1005                 8.66
1            7        0            0                0                  NaN
2           16       24           18              206                11.44
3           16       68           49              474                 9.67
```

```
4           12       71           54              420                  7.78

   Receiving TDs  Fumbles Lost  Total TD  Fantasy Points  PPR Fantasy Points
0              4             0        19             355               471.2
1              0             2         7             416               415.7
2              2             3        18             277               294.6
3              3             2        19             266               314.8
4              2             2        14             258               311.7

[5 rows x 24 columns]
```

[4]:
```
print("2020 Data Preview:\n", df_2020.head())
```

```
2020 Data Preview:
    Rank           Player Team Positon  Age  Games Played  Passing Completions
\
0     1  Derrick Henry*+  TEN      RB   26            16                    0
1     2    Alvin Kamara*  NOR      RB   25            15                    0
2     3     Dalvin Cook*  MIN      RB   25            14                    0
3     4  Davante Adams*+  GNB      WR   28            14                    0
4     5   Travis Kelce*+  KAN      TE   31            15                    1

   Passing Attempts  Passing Yards  Passing TD  …  Rushing TD  Targets   \
0                 0              0           0  …          17        31
1                 0              0           0  …          16       107
2                 0              0           0  …          16        54
3                 0              0           0  …           0       149
4                 2              4           0  …           0       145

   Receptions  Receiving Yards  Yards Per Reception  Receiving TD  \
0          19              114                 6.00             0
1          83              756                 9.11             5
2          44              361                 8.20             1
3         115             1374                11.95            18
4         105             1416                13.49            11

   Fumles Lost  Total TD  Fantasy Points  PPR Points
0            2        17             314       333.1
1            0        21             295       377.8
2            3        17             294       337.8
3            1        18             243       358.4
4            1        11             208       312.8

[5 rows x 24 columns]
```

[5]:
```
print("2021 Data Preview:\n", df_2021.head())
```

```
2021 Data Preview:
```

```
    Rank             Player Team Position  Age  Games Played  \
0     1  Jonathan Taylor*+  IND       RB   22            17
1     2      Cooper Kupp*+  LAR       WR   28            17
2     3     Deebo Samuel*+  SFO       WR   25            16
3     4         Josh Allen  BUF       QB   25            17
4     5      Austin Ekeler  LAC       RB   26            16

   Passing Completions  Passing Attempts  Passsing Yards  Passing TDs  …  \
0                    0                 0               0            0  …
1                    0                 1               0            0  …
2                    1                 2              24            1  …
3                  409               646            4407           36  …
4                    0                 0               0            0  …

   Rushing TDs  Target  Receptions  Receiving Yards  Yards Per Reception  \
0           18      51          40              360                 9.00
1            0     191         145             1947                13.43
2            8     121          77             1405                18.25
3            6       0           0                0                  NaN
4           12      94          70              647                 9.24

   Receiving Yards.1  Fumbles Lost  Total TDs  Fantasy Points  PPR Points
0                  2             2         20             333       373.1
1                 16             0         16             295       439.5
2                  6             2         14             262       339.0
3                  0             3          6             403       402.6
4                  8             3         20             274       343.8

[5 rows x 24 columns]
```

[6]:
```
print("2022 Data Preview:\n", df_2022.head())
```

```
2022 Data Preview:
    Rank               Player Team Positon  Age  Games Played  \
0     1     Patrick Mahomes*+  KAN       QB   27            17
1     2         Josh Jacobs*+  LVR       RB   24            17
2     3  Christian McCaffrey*  2TM       RB   26            17
3     4        Derrick Henry*  TEN       RB   28            16
4     5    Justin Jefferson*+  MIN       WR   23            17

   Passing Completions  Passing Attempts  Passing Yards  Passing Touchdowns  \
0                  435               648           5250                  41
1                    0                 0              0                   0
2                    1                 1             34                   1
3                    2                 2              4                   1
4                    2                 2             34                   0

   …  Rushing TD  Targets  Receptions  Receiving Yards  \
```

```
0  …           4        1           1               6
1  …          12       64          53             400
2  …           8      108          85             741
3  …          13       41          33             398
4  …           1      184         128            1809

   Yards per Receptions  Reciving Touchdowns  Fumbles Lost  Total TD2  \
0                  6.00                    0             5          4
1                  7.55                    0             3         12
2                  8.72                    5             1         13
3                 12.06                    0             6         13
4                 14.13                    8             0          9

   Fantasy Points  PPR Fantasy Points
0             416               417.4
1             275               328.3
2             271               356.4
3             270               302.8
4             241               368.7

[5 rows x 24 columns]
```

[7]:
```
# Check for missing values and data types for each year
print("2019 Data Information:")
print(df_2019.info())
```

```
2019 Data Information:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 24 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Rank                 200 non-null    int64
 1   Player               200 non-null    object
 2   Team                 200 non-null    object
 3   Position             200 non-null    object
 4   Age                  200 non-null    int64
 5   Games Played         200 non-null    int64
 6   Passing Completion   200 non-null    int64
 7   Passing Attempts     200 non-null    int64
 8   Passing Yards        200 non-null    int64
 9   Passing TDs          200 non-null    int64
 10  Interceptions        200 non-null    int64
 11  Rushing Attempts     200 non-null    int64
 12  Rushing Yards        200 non-null    int64
 13  Yards per Attempt    156 non-null    float64
 14  Rushing TDs          200 non-null    int64
 15  Targets              200 non-null    int64
```

```
 16  Recepotions           200 non-null    int64
 17  Receiving Yards       200 non-null    int64
 18  Yards per Reception   167 non-null    float64
 19  Receiving TDs         200 non-null    int64
 20  Fumbles Lost          200 non-null    int64
 21  Total TD              200 non-null    int64
 22  Fantasy Points        200 non-null    int64
 23  PPR Fantasy Points    200 non-null    float64
dtypes: float64(3), int64(18), object(3)
memory usage: 37.6+ KB
None
```

[8]:
```
print("\n2020 Data Information:")
print(df_2020.info())
```

```
2020 Data Information:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 24 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Rank                  200 non-null    int64
 1   Player                200 non-null    object
 2   Team                  200 non-null    object
 3   Positon               200 non-null    object
 4   Age                   200 non-null    int64
 5   Games Played          200 non-null    int64
 6   Passing Completions   200 non-null    int64
 7   Passing Attempts      200 non-null    int64
 8   Passing Yards         200 non-null    int64
 9   Passing TD            200 non-null    int64
 10  Interceptions         200 non-null    int64
 11  Rushing Attempts      200 non-null    int64
 12  Rushing Yards         200 non-null    int64
 13  Yards Per Attempt     155 non-null    float64
 14  Rushing TD            200 non-null    int64
 15  Targets               200 non-null    int64
 16  Receptions            200 non-null    int64
 17  Receiving Yards       200 non-null    int64
 18  Yards Per Reception   172 non-null    float64
 19  Receiving TD          200 non-null    int64
 20  Fumles Lost           200 non-null    int64
 21  Total TD              200 non-null    int64
 22  Fantasy Points        200 non-null    int64
 23  PPR Points            200 non-null    float64
dtypes: float64(3), int64(18), object(3)
memory usage: 37.6+ KB
```

```
None
```

[9]:
```
print("\n2021 Data Information:")
print(df_2021.info())
```

```
2021 Data Information:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 24 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Rank                  200 non-null    int64
 1   Player                200 non-null    object
 2   Team                  200 non-null    object
 3   Position              200 non-null    object
 4   Age                   200 non-null    int64
 5   Games Played          200 non-null    int64
 6   Passing Completions   200 non-null    int64
 7   Passing Attempts      200 non-null    int64
 8   Passsing Yards        200 non-null    int64
 9   Passing TDs           200 non-null    int64
 10  Interceptions         200 non-null    int64
 11  Rushing Attempts      200 non-null    int64
 12  Rushing Yards         200 non-null    int64
 13  Yards Per Attempt     162 non-null    float64
 14  Rushing TDs           200 non-null    int64
 15  Target                200 non-null    int64
 16  Receptions            200 non-null    int64
 17  Receiving Yards       200 non-null    int64
 18  Yards Per Reception   164 non-null    float64
 19  Receiving Yards.1     200 non-null    int64
 20  Fumbles Lost          200 non-null    int64
 21  Total TDs             200 non-null    int64
 22  Fantasy Points        200 non-null    int64
 23  PPR Points            200 non-null    float64
dtypes: float64(3), int64(18), object(3)
memory usage: 37.6+ KB
None
```

[10]:
```
print("\n2022 Data Information:")
print(df_2022.info())
```

```
2022 Data Information:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 24 columns):
```

```
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Rank                  200 non-null    int64
 1   Player                200 non-null    object
 2   Team                  200 non-null    object
 3   Positon               200 non-null    object
 4   Age                   200 non-null    int64
 5   Games Played          200 non-null    int64
 6   Passing Completions   200 non-null    int64
 7   Passing Attempts      200 non-null    int64
 8   Passing Yards         200 non-null    int64
 9   Passing Touchdowns    200 non-null    int64
 10  Interceptions         200 non-null    int64
 11  Rushing Attempts      200 non-null    int64
 12  Rushing Yards         200 non-null    int64
 13  Yards per Attempt     157 non-null    float64
 14  Rushing TD            200 non-null    int64
 15  Targets               200 non-null    int64
 16  Receptions            200 non-null    int64
 17  Receiving Yards       200 non-null    int64
 18  Yards per Receptions  166 non-null    float64
 19  Reciving Touchdowns   200 non-null    int64
 20  Fumbles Lost          200 non-null    int64
 21  Total TD2             200 non-null    int64
 22  Fantasy Points        200 non-null    int64
 23  PPR Fantasy Points    200 non-null    float64
dtypes: float64(3), int64(18), object(3)
memory usage: 37.6+ KB
None
```

[11]:
```
# Basic statistics for numerical columns (mean, min, max, etc.)
print("\n2019 Summary Statistics:")
print(df_2019.describe())
```

```
2019 Summary Statistics:
             Rank         Age  Games Played  Passing Completion  \
count  200.000000  200.000000    200.000000          200.000000
mean   100.500000   26.330000     14.105000           51.680000
std     57.879185    3.846437      2.460538          115.946174
min      1.000000   21.000000      3.000000            0.000000
25%     50.750000   24.000000     13.000000            0.000000
50%    100.500000   25.500000     15.000000            0.000000
75%    150.250000   28.000000     16.000000            0.000000
max    200.000000   42.000000     17.000000          408.000000

       Passing Attempts  Passing Yards   Passing TDs  Interceptions  \
count        200.000000      200.000000   200.000000     200.000000
```

```
mean          80.795000      592.350000     3.775000       1.755000
std          181.237935     1334.654668     8.575082       4.454682
min            0.000000        0.000000     0.000000       0.000000
25%            0.000000        0.000000     0.000000       0.000000
50%            0.000000        0.000000     0.000000       0.000000
75%            1.000000        0.000000     0.000000       0.000000
max          626.000000     5109.000000    36.000000      30.000000

       Rushing Attempts  Rushing Yards  …  Rushing TDs     Targets  \
count        200.000000     200.000000  …   200.000000  200.000000
mean          55.970000     246.340000  …     1.960000   59.705000
std           81.342066     366.605212  …     3.163612   44.292404
min            0.000000     -12.000000  …     0.000000    0.000000
25%            1.000000       0.000000  …     0.000000   21.750000
50%            9.000000      40.500000  …     0.000000   56.500000
75%           82.250000     374.250000  …     3.000000   90.250000
max          303.000000    1540.000000  …    16.000000  185.000000

       Recepotions  Receiving Yards  Yards per Reception  Receiving TDs  \
count   200.000000       200.000000           167.000000     200.000000
mean     40.120000       471.120000            11.116766       3.055000
std      29.785086       388.879547             3.683657       2.760448
min       0.000000        -4.000000            -4.000000       0.000000
25%      14.750000       118.750000             8.325000       0.000000
50%      39.000000       424.500000            11.180000       3.000000
75%      59.000000       715.250000            13.720000       5.000000
max     149.000000      1725.000000            20.690000      11.000000

       Fumbles Lost    Total TD  Fantasy Points  PPR Fantasy Points
count    200.000000  200.000000      200.000000          200.000000
mean       1.090000    5.040000      135.945000          176.032000
std        1.585709    3.275982       70.621448           73.096789
min        0.000000    0.000000       55.000000           58.100000
25%        0.000000    3.000000       79.750000          113.575000
50%        1.000000    5.000000      118.500000          164.550000
75%        2.000000    7.000000      168.000000          225.500000
max       11.000000   19.000000      416.000000          471.200000

[8 rows x 21 columns]
```

[12]:
```
print("\n2020 Summary Statistics:")
print(df_2020.describe())
```

```
2020 Summary Statistics:
             Rank         Age  Games Played  Passing Completions  \
count  200.000000  200.000000    200.000000            200.00000
mean   100.500000   26.345000     14.040000             54.08000
```

```
std     57.879185    3.932339      2.598492            118.08405
min      1.000000   21.000000      3.000000              0.00000
25%     50.750000   24.000000     13.000000              0.00000
50%    100.500000   25.000000     15.000000              0.00000
75%    150.250000   28.000000     16.000000              0.25000
max    200.000000   43.000000     16.000000            407.00000

       Passing Attempts  Passing Yards  Passing TD  Interceptions  \
count         200.00000     200.000000  200.000000     200.000000
mean           82.13500     601.650000    4.085000       1.715000
std           178.39076    1325.122438    9.709229       3.758244
min             0.00000       0.000000    0.000000       0.000000
25%             0.00000       0.000000    0.000000       0.000000
50%             0.00000       0.000000    0.000000       0.000000
75%             1.00000       1.000000    0.000000       0.000000
max           626.00000    4823.000000   48.000000      15.000000

       Rushing Attempts  Rushing Yards  …  Rushing TD    Targets   \
count       200.000000     200.000000  …  200.000000  200.000000
mean         54.825000     248.605000  …    2.335000   59.025000
std          74.243687     351.093982  …    3.442678   44.046508
min           0.000000      -8.000000  …    0.000000    0.000000
25%           1.000000       0.000000  …    0.000000   19.000000
50%          11.500000      47.500000  …    1.000000   59.000000
75%          97.000000     429.500000  …    3.000000   92.250000
max         378.000000    2027.000000  …   17.000000  166.000000

       Receptions  Receiving Yards  Yards Per Reception  Receiving TD  \
count   200.00000       200.000000           172.000000     200.00000
mean     40.38000       459.670000            10.541744       3.22000
std      29.97764       385.912234             3.943376       3.32821
min       0.00000        -6.000000            -6.000000       0.00000
25%      16.00000       122.250000             7.740000       0.00000
50%      38.00000       418.000000            10.760000       3.00000
75%      59.00000       723.750000            13.222500       5.00000
max     127.00000      1535.000000            20.910000      18.00000

       Fumles Lost    Total TD  Fantasy Points  PPR Points
count    200.00000  200.000000      200.000000  200.000000
mean       0.93000    5.580000      140.285000  180.623500
std        1.39457    3.631742       74.926052   75.091079
min        0.00000    0.000000       63.000000   64.300000
25%        0.00000    3.000000       86.000000  126.875000
50%        0.00000    5.000000      116.000000  164.200000
75%        1.00000    7.000000      166.000000  223.725000
max        8.00000   21.000000      395.000000  396.100000

[8 rows x 21 columns]
```

[13]:
```
print("\n2021 Summary Statistics:")
print(df_2021.describe())
```

```
2021 Summary Statistics:
              Rank         Age  Games Played  Passing Completions  \
count   200.000000  200.000000    200.000000            200.00000
mean    100.500000   26.270000     14.570000             55.75000
std      57.879185    3.663908      2.852329            123.65529
min       1.000000   21.000000      6.000000              0.00000
25%      50.750000   24.000000     13.000000              0.00000
50%     100.500000   26.000000     16.000000              0.00000
75%     150.250000   28.000000     17.000000              0.25000
max     200.000000   44.000000     17.000000            485.00000

        Passing Attempts  Passsing Yards  Passing TDs  Interceptions  \
count         200.000000      200.000000   200.000000     200.000000
mean           85.710000      614.010000     3.935000       1.945000
std           188.009798     1365.233155     9.389975       4.304021
min             0.000000        0.000000     0.000000       0.000000
25%             0.000000        0.000000     0.000000       0.000000
50%             0.000000        0.000000     0.000000       0.000000
75%             1.000000        1.000000     0.000000       0.000000
max           719.000000     5316.000000    43.000000      17.000000

        Rushing Attempts   Rushing Yards  …  Rushing TDs      Target  \
count         200.000000     200.000000  …    200.00000  200.000000
mean           59.350000     263.215000  …      2.19000   60.150000
std            77.035184     342.780954  …      3.20237   46.423737
min             0.000000       0.000000  …      0.00000    0.000000
25%             1.000000       5.000000  …      0.00000   20.000000
50%            17.500000      78.500000  …      1.00000   57.500000
75%           104.250000     435.250000  …      3.00000   92.250000
max           332.000000    1811.000000  …     18.00000  191.000000

        Receptions  Receiving Yards  Yards Per Reception  Receiving Yards.1  \
count   200.00000       200.000000           164.000000         200.000000
mean     41.16500       465.720000            10.696951           3.015000
std      30.97718       401.140274             3.520453           3.224401
min       0.00000        -4.000000            -4.000000           0.000000
25%      18.00000       128.750000             7.977500           0.000000
50%      41.00000       430.500000            10.690000           2.000000
75%      61.00000       705.000000            13.122500           5.000000
max     145.00000      1947.000000            19.540000          16.000000

        Fumbles Lost   Total TDs  Fantasy Points  PPR Points
count    200.000000  200.000000      200.000000  200.000000
```

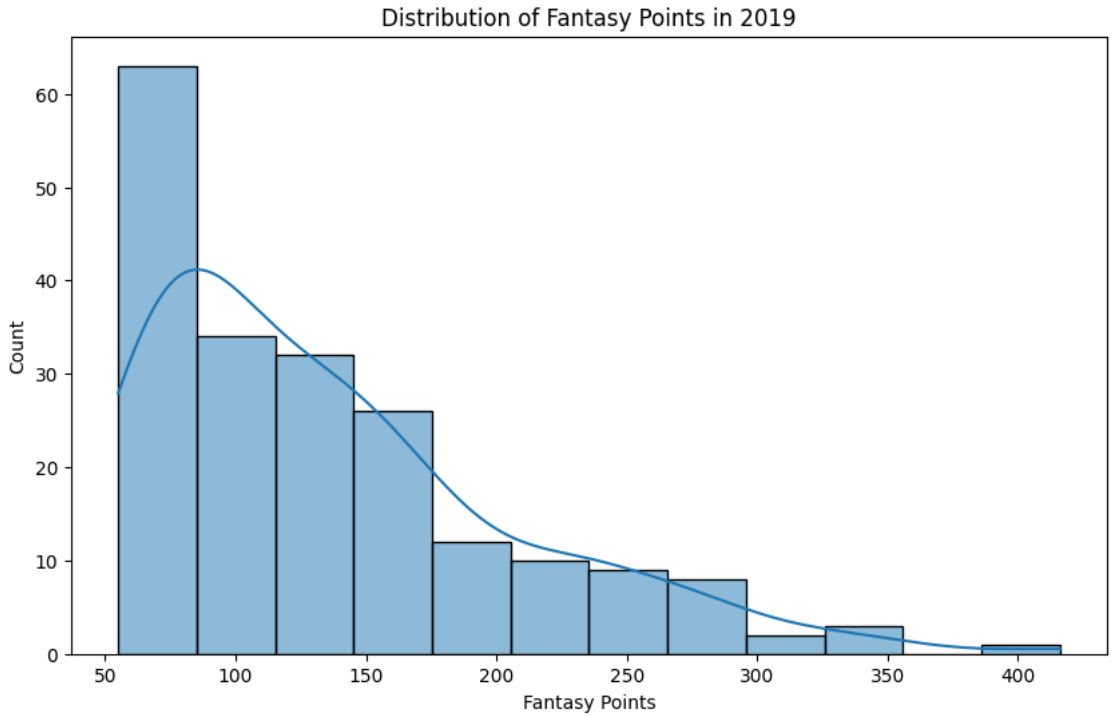
```
mean       0.960000    5.220000      139.775000  180.869000
std        1.306497    3.691849       73.146335   75.794843
min        0.000000    0.000000       59.000000   62.500000
25%        0.000000    3.000000       85.000000  121.700000
50%        1.000000    5.000000      116.500000  164.200000
75%        1.000000    7.000000      172.000000  227.350000
max        6.000000   20.000000      403.000000  439.500000

[8 rows x 21 columns]
```

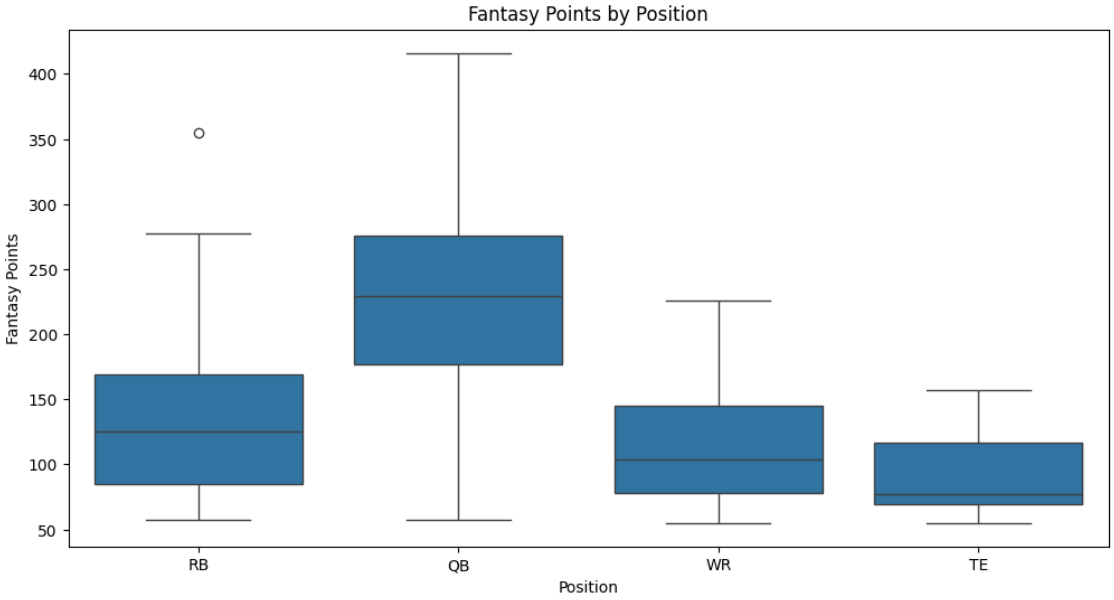
[14]:
```
print("\n2022 Summary Statistics:")
print(df_2022.describe())
```

```
2022 Summary Statistics:
             Rank         Age  Games Played  Passing Completions  \
count  200.000000  200.000000    200.000000           200.000000
mean   100.500000   26.385000     14.585000            50.680000
std     57.879185    3.438881      2.760594           114.872346
min      1.000000   21.000000      6.000000             0.000000
25%     50.750000   24.000000     13.000000             0.000000
50%    100.500000   26.000000     16.000000             0.000000
75%    150.250000   28.000000     17.000000             0.000000
max    200.000000   45.000000     17.000000           490.000000

       Passing Attempts  Passing Yards  Passing Touchdowns  Interceptions  \
count        200.000000     200.000000          200.000000     200.000000
mean          78.235000     558.965000            3.445000       1.675000
std          175.443014    1256.461792            8.089759       3.744259
min            0.000000       0.000000            0.000000       0.000000
25%            0.000000       0.000000            0.000000       0.000000
50%            0.000000       0.000000            0.000000       0.000000
75%            1.000000       0.000000            0.000000       0.000000
max          733.000000    5250.000000           41.000000      15.000000

       Rushing Attempts  Rushing Yards   …  Rushing TD     Targets  \
count        200.000000     200.000000  …  200.000000  200.000000
mean          60.630000     274.515000  …    2.110000   59.965000
std           83.747845     388.655432  …    3.210833   45.957089
min            0.000000     -15.000000  …    0.000000    0.000000
25%            1.000000       0.000000  …    0.000000   18.000000
50%           10.000000      53.500000  …    1.000000   59.000000
75%           95.000000     462.250000  …    3.000000   92.250000
max          349.000000    1653.000000  …   17.000000  184.000000

       Receptions  Receiving Yards  Yards per Receptions  Reciving Touchdowns  \
count  200.000000        200.00000            166.000000           200.000000
mean    40.730000        456.14500             10.348193             2.765000
```

```
std     30.634613        396.47656              3.571122             2.867208
min      0.000000        -10.00000             -5.000000             0.000000
25%     15.750000         95.75000              7.682500             0.000000
50%     40.000000        423.50000             10.585000             2.000000
75%     60.250000        710.75000             12.872500             4.000000
max    128.000000       1809.00000             18.080000            14.000000

       Fumbles Lost   Total TD2  Fantasy Points  PPR Fantasy Points
count     200.00000  200.000000       200.00000          200.000000
mean        2.11000    4.905000       134.09000          174.766000
std         2.79229    3.319801        71.48962           74.628082
min         0.00000    0.000000        57.00000           56.600000
25%         0.00000    3.000000        79.75000          115.100000
50%         1.00000    4.000000       115.00000          162.600000
75%         3.00000    6.000000       165.75000          219.550000
max        16.00000   18.000000       416.00000          417.400000

[8 rows x 21 columns]
```

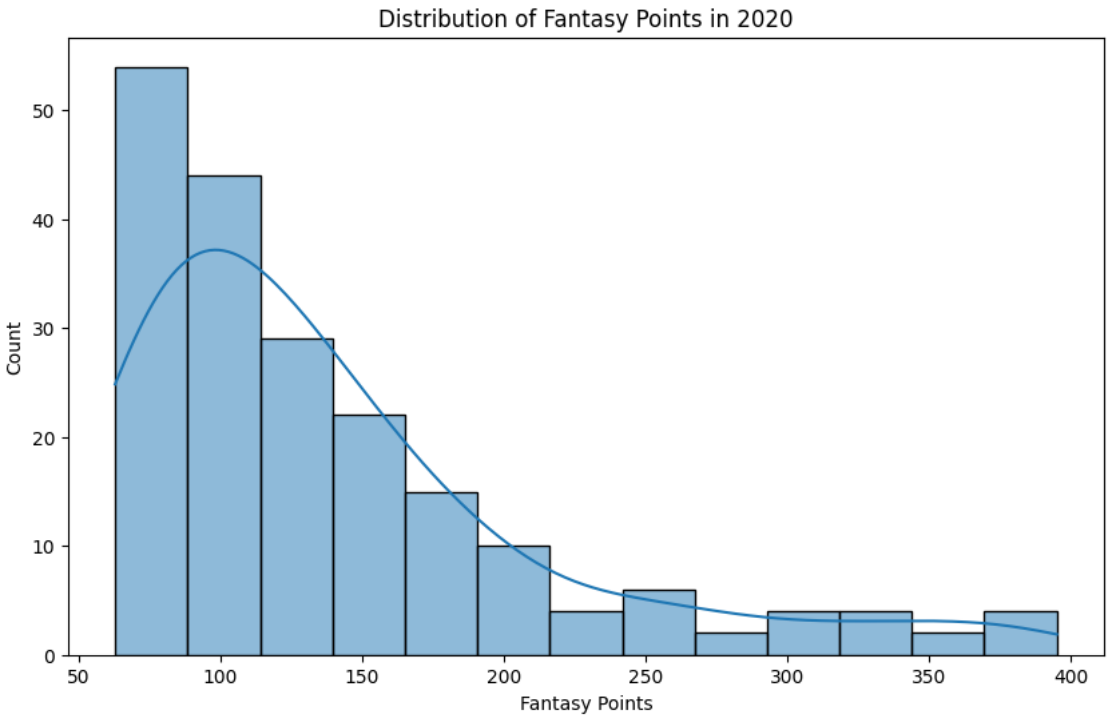
[15]:
```python
import matplotlib.pyplot as plt
import seaborn as sns

# Plot distribution of fantasy points
plt.figure(figsize=(10, 6))
sns.histplot(df_2019['Fantasy Points'], kde=True)
plt.title('Distribution of Fantasy Points in 2019')
plt.show()

# Boxplot to compare fantasy points across positions
plt.figure(figsize=(12, 6))
sns.boxplot(x='Position', y='Fantasy Points', data=df_2019)
plt.title('Fantasy Points by Position')
plt.show()
```

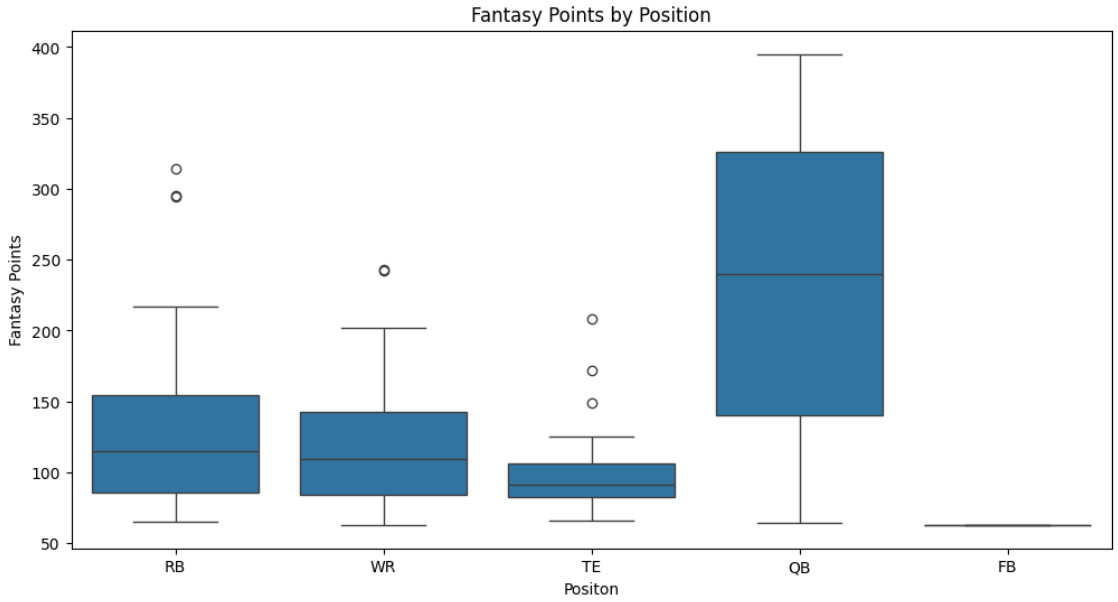




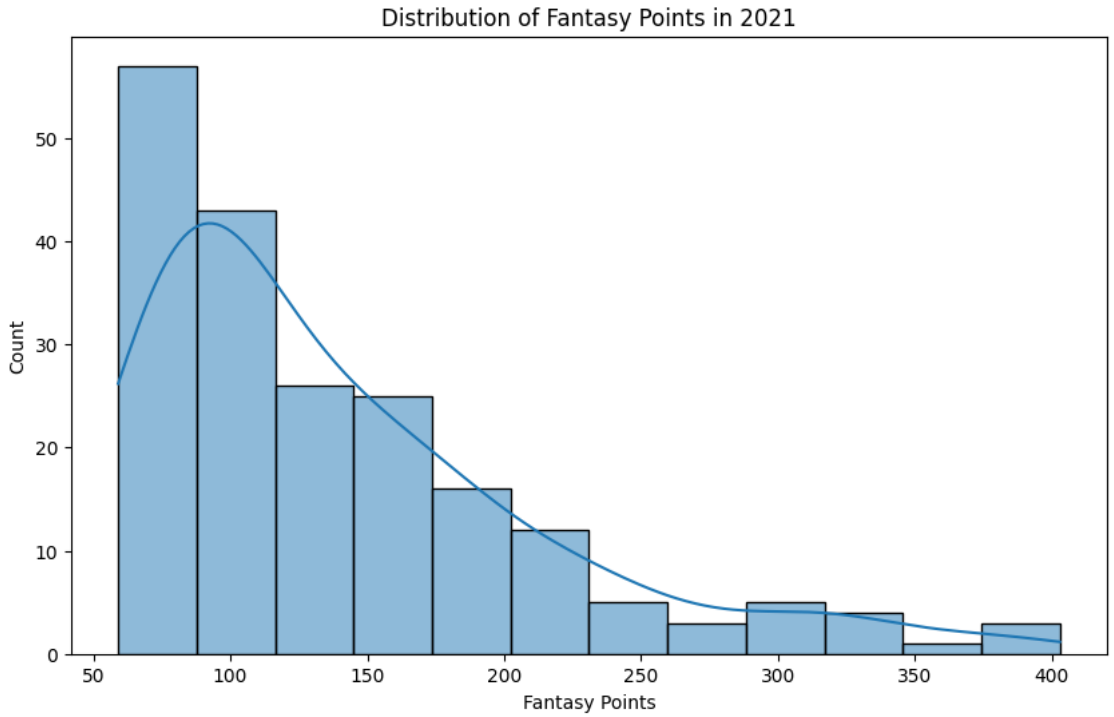
[16]:
```
import matplotlib.pyplot as plt
import seaborn as sns

# Plot distribution of fantasy points
```

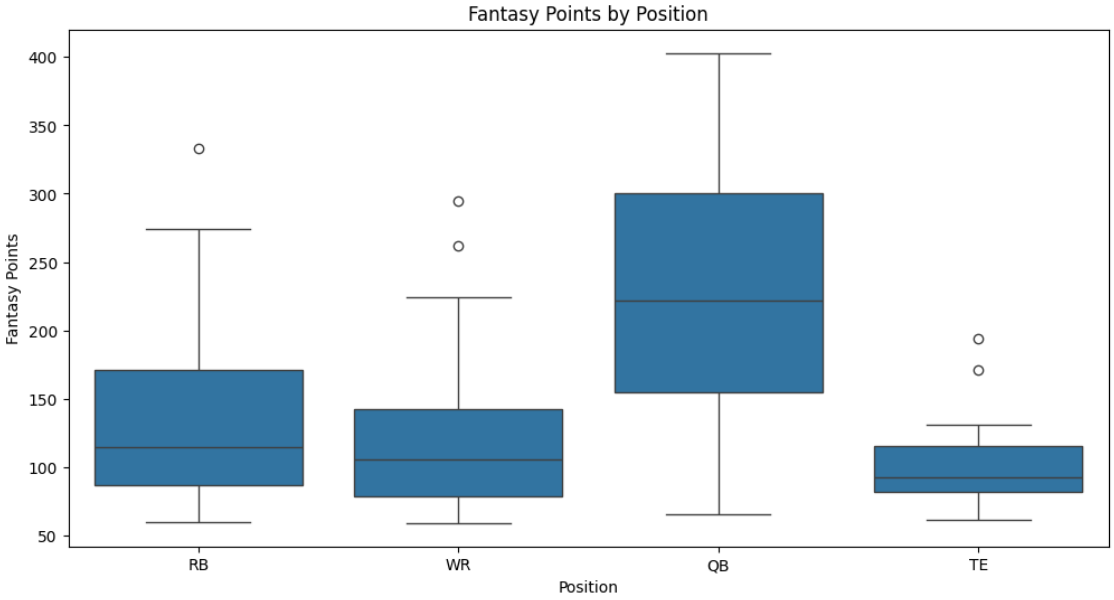
```
plt.figure(figsize=(10, 6))
sns.histplot(df_2020['Fantasy Points'], kde=True)
plt.title('Distribution of Fantasy Points in 2020')
plt.show()

# Boxplot to compare fantasy points across positions
plt.figure(figsize=(12, 6))
sns.boxplot(x='Positon', y='Fantasy Points', data=df_2020)
plt.title('Fantasy Points by Position')
plt.show()
```





[17]:
```
import matplotlib.pyplot as plt
import seaborn as sns

# Plot distribution of fantasy points
plt.figure(figsize=(10, 6))
sns.histplot(df_2021['Fantasy Points'], kde=True)
plt.title('Distribution of Fantasy Points in 2021')
plt.show()

# Boxplot to compare fantasy points across positions
plt.figure(figsize=(12, 6))
sns.boxplot(x='Position', y='Fantasy Points', data=df_2021)
plt.title('Fantasy Points by Position')
plt.show()
```





[18]:
```
############ START OF WEEK ############
```

[19]:
```
# Import necessary libraries
import re
import pandas as pd
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
import seaborn as sns
```

[20]:
```
# Load data from CSV files
df1 = pd.read_csv('FFRank 2019.csv')
df2 = pd.read_csv('FFRank 2020.csv')
df3 = pd.read_csv('FFRank 2021.csv')
df4 = pd.read_csv('FFRank 2022.csv')
```

[21]:
```
# Concatenate all dataframes into a single dataframe
combined_df = pd.concat([df1, df2, df3, df4], ignore_index=True)

# Clean player names by removing special characters
combined_df['Player'] = combined_df['Player'].str.replace(r'[^a-zA-Z.\s]', '',␣
 ↪regex=True)
```

[22]:
```
# Group by 'Player' and aggregate relevant numerical columns
aggregated_df = combined_df.groupby('Player').agg({
    'Rushing Yards': 'sum',
    'Receiving Yards': 'sum',
    'Passing Yards': 'sum',
    'Total TD': 'sum',
    'Fantasy Points': 'sum',
    'Games Played': 'sum',
    'Position': 'first',  # Get the first non-null position
}).reset_index()
```

[23]:
```
# Step 1.1: Calculate Yards from Scrimmage and Total Yards
aggregated_df['Yards_from_Scrimmage'] = aggregated_df['Rushing Yards'] +␣
 ↪aggregated_df['Receiving Yards']
aggregated_df['Total_Yards'] = aggregated_df['Yards_from_Scrimmage'] +␣
 ↪aggregated_df['Passing Yards']
```

[24]:
```
# Step 1.2: Optional - Calculate averages for aggregated statistics
aggregated_df['Avg_TD'] = aggregated_df['Total TD'] / aggregated_df['Games␣
 ↪Played']
aggregated_df['Avg_Yards_from_Scrimmage'] =␣
 ↪aggregated_df['Yards_from_Scrimmage'] / aggregated_df['Games Played']
aggregated_df['Avg_Passing_Yards'] = aggregated_df['Passing Yards'] /␣
 ↪aggregated_df['Games Played']
aggregated_df['Avg_Total_Yards'] = aggregated_df['Total_Yards'] /␣
 ↪aggregated_df['Games Played']
```

[25]:
```
# Save the aggregated data to a new CSV file
aggregated_df.to_csv('aggregated_fantasy_data.csv', index=False)
```

[26]:
```
# Display the aggregated data
print(aggregated_df.head())
```

```
          Player  Rushing Yards  Receiving Yards  Passing Yards  Total TD  \
0     A.J. Brown           70.0             4491            0.0      21.0
1     A.J. Green            0.0             1371            0.0       2.0
2      AJ Dillon          803.0              519            0.0       0.0
3    Aaron Jones         2987.0             1615            0.0      30.0
4  Aaron Rodgers          433.0              -10         7994.0       4.0

   Fantasy Points  Games Played Position  Yards_from_Scrimmage  Total_Yards  \
0             673            60       WR                4561.0       4561.0
1             167            32       WR                1371.0       1371.0
2             292            34       RB                1322.0       1322.0
3             845            62       RB                4602.0       4602.0
4            1231            65       QB                 423.0       8417.0

     Avg_TD  Avg_Yards_from_Scrimmage  Avg_Passing_Yards  Avg_Total_Yards
0  0.350000                 76.016667           0.000000        76.016667
1  0.062500                 42.843750           0.000000        42.843750
2  0.000000                 38.882353           0.000000        38.882353
3  0.483871                 74.225806           0.000000        74.225806
4  0.061538                  6.507692         122.984615       129.492308
```

[27]:
```
# ============== Step 2: Split Data (Train/Test) ==============

# Features: Yards_from_Scrimmage, Passing Yards, TD, etc.
X = aggregated_df[['Yards_from_Scrimmage', 'Passing Yards', 'Total TD',␣
 ↪'Total_Yards']]
y = aggregated_df['Fantasy Points']  # Fantasy Points as the target

# Split the data into 80% training and 20% test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=42)
```

[28]:
```
# Confirm the split sizes
print("Train Feature Set Shape:", X_train.shape)
print("Test Feature Set Shape:", X_test.shape)
print("Train Target Set Shape:", y_train.shape)
print("Test Target Set Shape:", y_test.shape)
```

```
Train Feature Set Shape: (288, 4)
Test Feature Set Shape: (72, 4)
Train Target Set Shape: (288,)
```

```
Test Target Set Shape: (72,)
```

[29]:
```
# Overview of the data
print("\nData Overview:")
print(aggregated_df.head())
```

```
Data Overview:
          Player  Rushing Yards  Receiving Yards  Passing Yards  Total TD  \
0     A.J. Brown           70.0             4491            0.0      21.0
1     A.J. Green            0.0             1371            0.0       2.0
2      AJ Dillon          803.0              519            0.0       0.0
3    Aaron Jones         2987.0             1615            0.0      30.0
4  Aaron Rodgers          433.0              -10         7994.0       4.0

   Fantasy Points  Games Played Position  Yards_from_Scrimmage  Total_Yards  \
0             673            60       WR                4561.0       4561.0
1             167            32       WR                1371.0       1371.0
2             292            34       RB                1322.0       1322.0
3             845            62       RB                4602.0       4602.0
4            1231            65       QB                 423.0       8417.0

     Avg_TD  Avg_Yards_from_Scrimmage  Avg_Passing_Yards  Avg_Total_Yards
0  0.350000                 76.016667           0.000000        76.016667
1  0.062500                 42.843750           0.000000        42.843750
2  0.000000                 38.882353           0.000000        38.882353
3  0.483871                 74.225806           0.000000        74.225806
4  0.061538                  6.507692         122.984615       129.492308
```

[30]:
```
# Check for missing values
print("\nMissing Values:")
print(aggregated_df.isnull().sum())
```

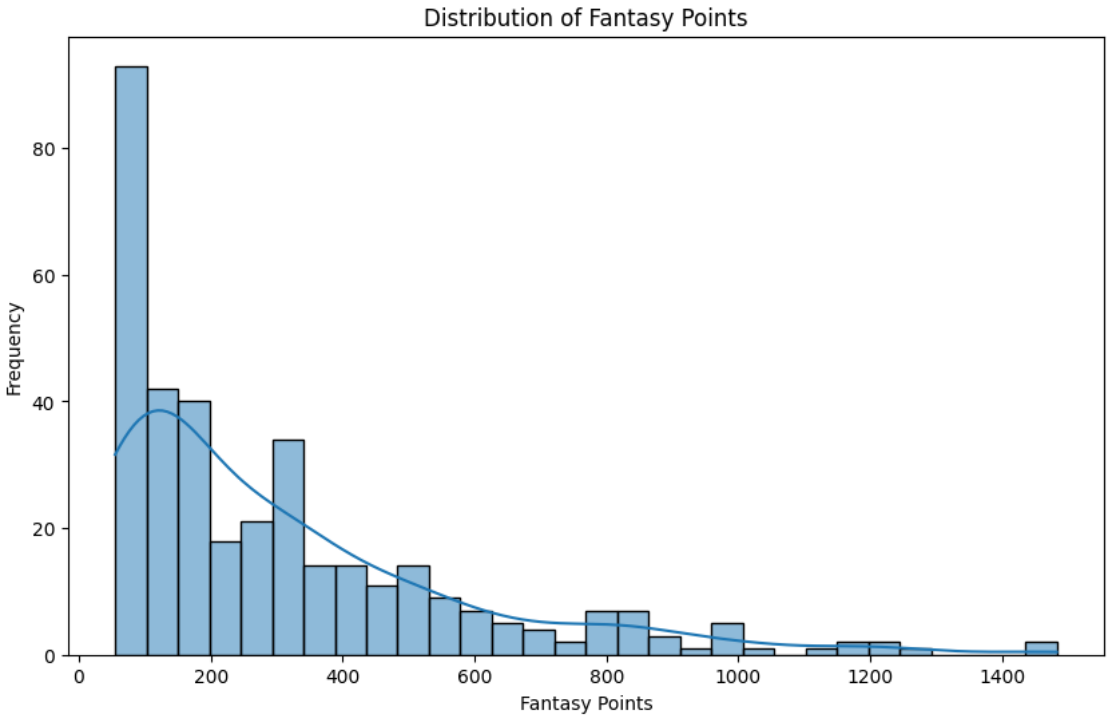
```
Missing Values:
Player                     0
Rushing Yards              0
Receiving Yards            0
Passing Yards              0
Total TD                   0
Fantasy Points             0
Games Played               0
Position                  68
Yards_from_Scrimmage       0
Total_Yards                0
Avg_TD                     0
Avg_Yards_from_Scrimmage   0
Avg_Passing_Yards          0
```

```
Avg_Total_Yards              0
dtype: int64
```

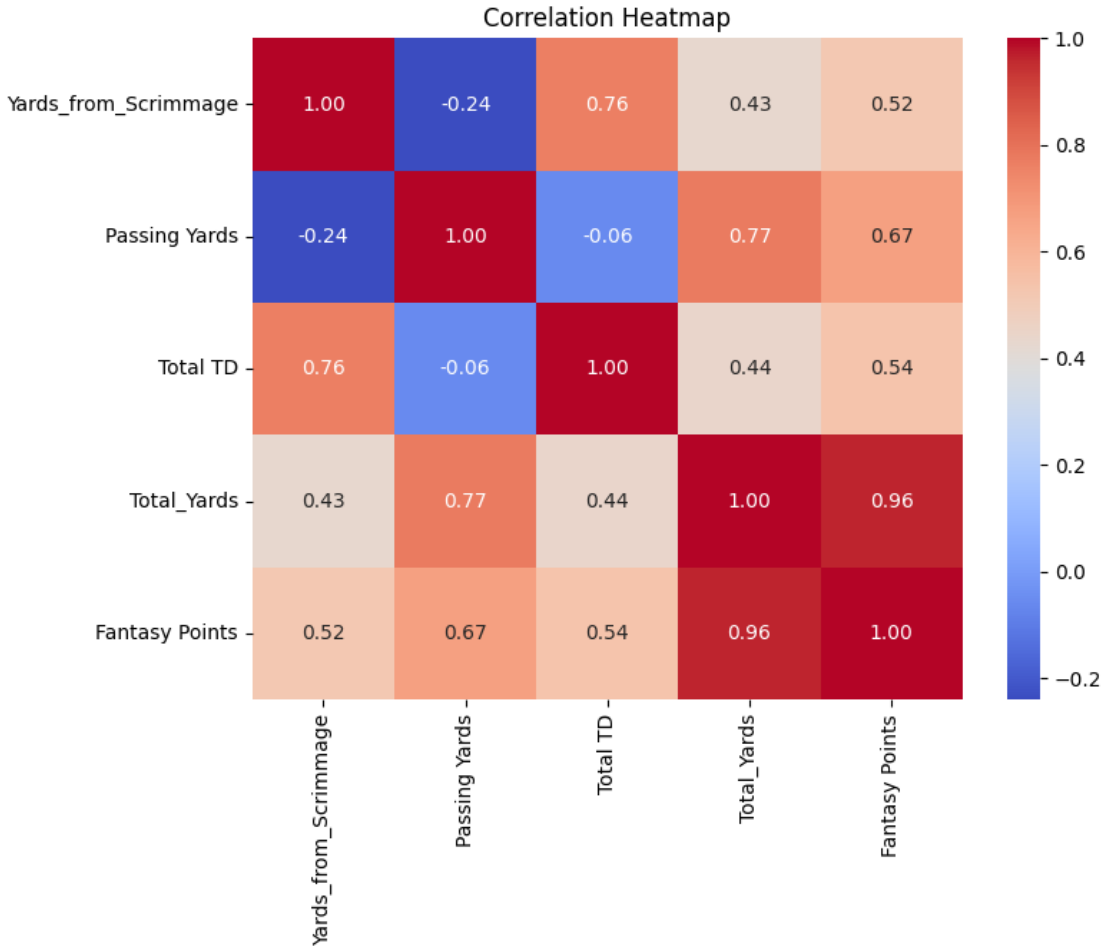
[31]:
```
# Summary statistics of the numerical columns
print("\nSummary Statistics:")
print(aggregated_df.describe())
```

```
Summary Statistics:
       Rushing Yards  Receiving Yards  Passing Yards   Total TD  \
count     360.000000       360.000000     360.000000  360.00000
mean      421.200000      1029.252778     644.786111    5.90000
std       744.057489      1168.082611    1843.346207    6.18075
min       -12.000000       -10.000000       0.000000    0.00000
25%         0.000000       200.500000       0.000000    0.00000
50%        37.500000       576.000000       0.000000    4.50000
75%       504.750000      1273.500000       0.000000    9.00000
max      4504.000000      5440.000000    9990.000000   35.00000

       Fantasy Points  Games Played  Yards_from_Scrimmage   Total_Yards  \
count      360.000000    360.000000            360.000000    360.000000
mean       305.608333     31.833333           1450.452778   2095.238889
std        271.371600     17.668725           1292.883314   1980.713934
min         55.000000      3.000000             -6.000000     12.000000
25%        100.500000     16.000000            487.250000    629.000000
50%        212.000000     30.000000           1017.500000   1430.500000
75%        422.000000     46.000000           2215.500000   2720.000000
max       1483.000000     66.000000           5440.000000  10903.000000

           Avg_TD  Avg_Yards_from_Scrimmage  Avg_Passing_Yards  \
count  360.000000                360.000000         360.000000
mean     0.175281                 42.503163          17.564874
std      0.163562                 22.568150          46.170247
min      0.000000                 -0.260870           0.000000
25%      0.000000                 28.183293           0.000000
50%      0.146687                 42.021739           0.000000
75%      0.279615                 57.836976           0.000000
max      0.766667                107.250000         251.882353

       Avg_Total_Yards
count       360.000000
mean         60.068037
std          38.432380
min           1.200000
25%          37.039299
50%          49.836439
75%          69.612931
max         251.882353
```

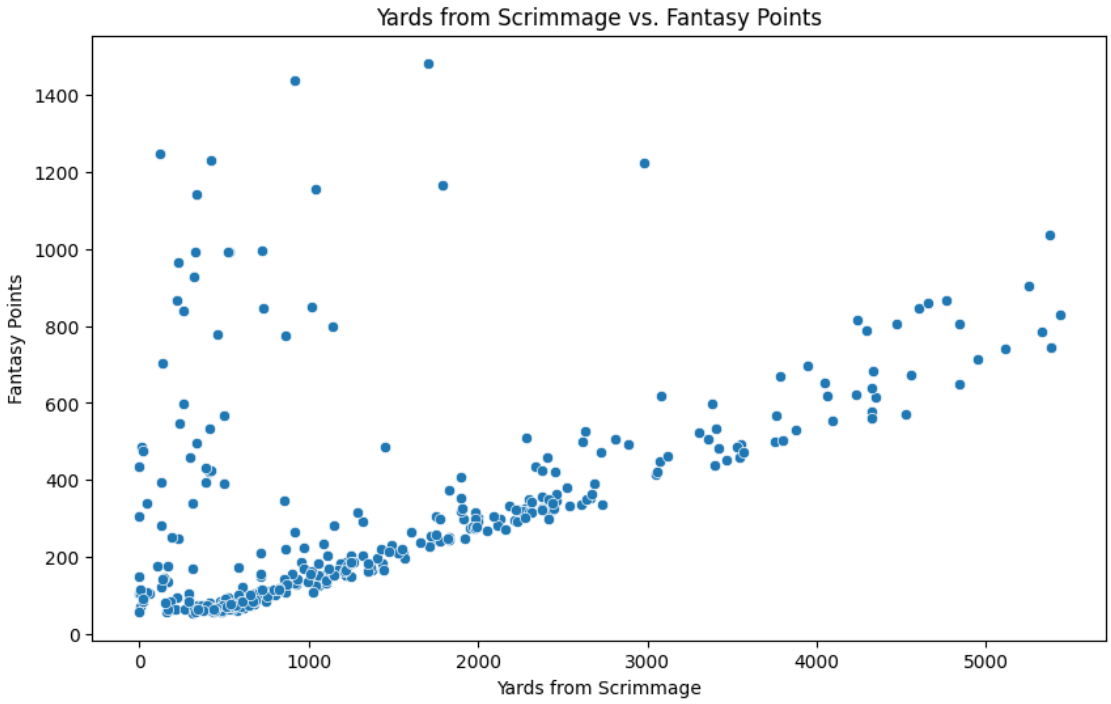
[32]:
```
# Visualize the distribution of fantasy points
plt.figure(figsize=(10, 6))
sns.histplot(aggregated_df['Fantasy Points'], bins=30, kde=True)
plt.title('Distribution of Fantasy Points')
plt.xlabel('Fantasy Points')
plt.ylabel('Frequency')
plt.show()
```



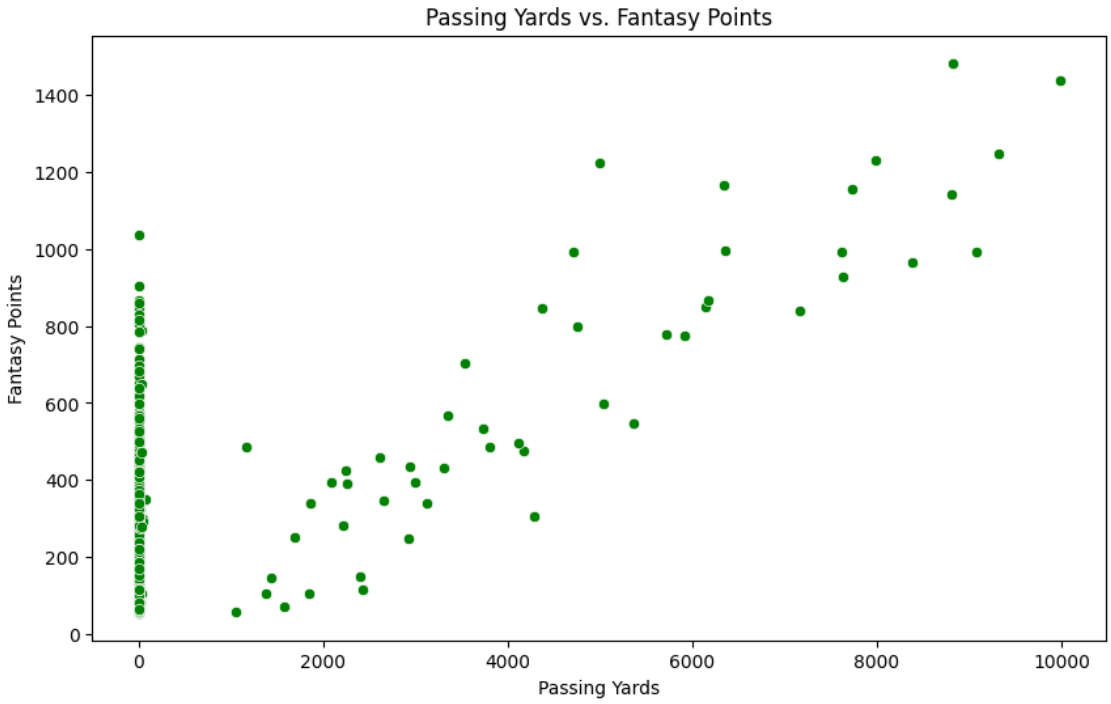
[33]:
```
# Correlation heatmap for numerical columns
plt.figure(figsize=(8, 6))
corr = aggregated_df[['Yards_from_Scrimmage', 'Passing Yards', 'Total TD',␣
↪'Total_Yards', 'Fantasy Points']].corr()
sns.heatmap(corr, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Heatmap')
plt.show()
```



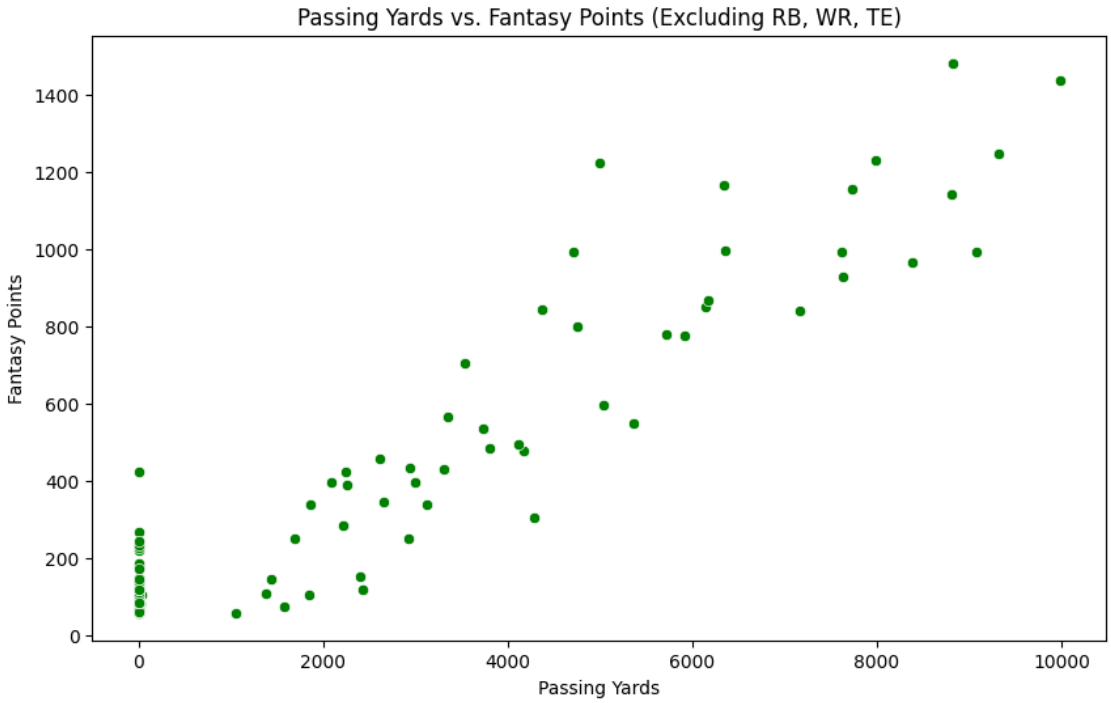
[34]:
```
# Scatterplot: Yards from Scrimmage vs. Fantasy Points
plt.figure(figsize=(10, 6))
sns.scatterplot(x=aggregated_df['Yards_from_Scrimmage'],␣
 ↪y=aggregated_df['Fantasy Points'])
plt.title('Yards from Scrimmage vs. Fantasy Points')
plt.xlabel('Yards from Scrimmage')
plt.ylabel('Fantasy Points')
plt.show()
```



[35]:
```
# Scatterplot: Passing Yards vs. Fantasy Points
plt.figure(figsize=(10, 6))
sns.scatterplot(x=aggregated_df['Passing Yards'], y=aggregated_df['Fantasy␣
↪Points'], color='green')
plt.title('Passing Yards vs. Fantasy Points')
plt.xlabel('Passing Yards')
plt.ylabel('Fantasy Points')
plt.show()
```



[36]:
```
# Filter out RB, WR, and TE players
filtered_data = aggregated_df[~aggregated_df['Position'].isin(['RB', 'WR',␣
↪'TE'])]
# Scatterplot: Passing Yards vs. Fantasy Points (Excluding RB, WR, TE)
plt.figure(figsize=(10, 6))
sns.scatterplot(x=filtered_data['Passing Yards'], y=filtered_data['Fantasy␣
↪Points'], color='green')
plt.title('Passing Yards vs. Fantasy Points (Excluding RB, WR, TE)')
plt.xlabel('Passing Yards')
plt.ylabel('Fantasy Points')
plt.show()
```



[37]:
```
# =============== Step 5: Save the Train/Test Data ===============

# Save the training and test sets to CSV for future use
X_train.to_csv('X_train.csv', index=False)
X_test.to_csv('X_test.csv', index=False)
y_train.to_csv('y_train.csv', index=False)
y_test.to_csv('y_test.csv', index=False)
```

[38]:
```
############### END OF WEEK ###############
```

[ ]:

[ ]:

[ ]:

[ ]:

[ ]: