Analysis of World Football Transfers in the 2021 Calendar Year

Christopher Spann

Background

- McDaniel College – Master's of Data Analytics
  - Fundamentals of Python Programming – Data Analytics Project

- Purpose: Choose a dataset of interest and establish research findings and conclusions
  - Data Prep/Cleaning
  - Data Munging
  - Data Visualization
  - Data Aggregation and Grouping

Introduction to Project and Data

- Goal: To better summarize statistics of football transfers that occurred in the 2021 calendar year using data cleaning, transformation, organization, and visualization techniques in Python
  - Focus on permanent transfers between two clubs with a fee greater than 0.
  - Attributes of interest
    - Club Destination
    - Player Position
    - Player Age

- Dataset - Football Transfers 2021 (Kaggle)
  - Scraped from transfermarkt site
  - 65,489 records
  - 11 variables

Data Cleaning & Preparation

- Removed all records in which the player retired, went without club, took a career break, was banned, or destination is unknown
- Created a clean transfer fee column by removing text, symbols, etc. and utilizing the unit (millions or thousands)
- Updated clean transfer fee value to 0 for records not applicable to research scope
- Removed all records in which transfer fee = 0 or equivalent
  - Loans, Drafts, Free Transfers, Null Values
- Final Dataset: 1,914 records

```
Unique Values of the transfer_fee column:
['€22.00m' '€8.50m' '€5.45m' '€4.50m' '€4.48m' '€4.00m' '€3.70m' '€3.65m'
 '€3.50m' '€3.36m' '€2.54m' '€2.00m' '€1.80m' '€1.67m' '€1.60m' '€1.59m'
 '€1.50m' '€1.36m' 'Loan fee:€1.00m' '€1.00m' '€822Th.' '€800Th.'
 '€760Th.' '€636Th.' '€590Th.' '€545Th.' '€500Th.' '€410Th.' '€400Th.'
 '€350Th.' '€330Th.' '€300Th.' '€286Th.' '€272Th.' '€250Th.' '€200Th.'
 'Loan fee:€200Th.' '€164Th.' '€135Th.' '€120Th.' '€100Th.' '€95Th.'
 '€55Th.' '€50Th.' '€40Th.' 'Loan fee:€25Th.' '€20Th.' '€12Th.' '€10Th.'
 '€1Th.' 'free transfer' 'loan transfer' '?' '-' '€21.30m' '€2.42m'
 '€850Th.' '€12.00m' '€4.10m' '€3.00m' '€2.21m' 'Loan fee:€500Th.'
 '€268Th.' '€169Th.' '€14.00m' '€2.50m' '€1.75m' 'Loan fee:€100Th.'
 '€70Th.' '€160Th.' '€1.91m' '€600Th.' '€150Th.' '€96Th.' '€22.50m'
 'Loan fee:€1.25m' '€110Th.' '€824Th.' '€490Th.' '€408Th.' '€11Th.'
 'Loan fee:€175Th.' 'Loan fee:€105Th.' '€455Th.' '€290Th.' '€206Th.'
 '€165Th.' 'Loan fee:€133Th.' '€80Th.' '€6.00m' '€22Th.' '€1.70m'
 'Loan fee:€1.50m' 'Loan fee:€550Th.' '€123Th.' '€30Th.' '€75Th.' '€1.32m'
 'Loan fee:€80Th.' '€314Th.' '€3.31m' '€65Th.' '€60Th.' '€975Th.'
 '€770Th.' '€909Th.' '€700Th.' '€409Th.' '€295Th.' 'Loan fee:€250Th.'
 'Loan fee:€150Th.' '€90Th.' nan '€220Th.' '€28Th.' '€16.70m' '€9.00m'
 'Loan fee:€3.93m' 'Loan fee:€3.50m' '€2.90m' '€2.27m' '€370Th.' '€227Th.'
 '€223Th.' '€170Th.' 'draft' 'Loan fee:€600Th.' '€471Th.' '€380Th.'
 '€130Th.' 'Loan fee:€3.00m' '€15Th.' '€10.00m' 'Loan fee:€1.20m'
 'Loan fee:€650Th.' '€345Th.' '€15.80m' '€5.00m' '€148Th.' '€11.00m'
 'Loan fee:€2.00m' '€282Th.' '€125Th.' 'Loan fee:€8Th.' '€23.30m' '€8.00m'
 '€508Th.' '€23.10m' '€13.00m' 'Loan fee:€2.30m' 'Loan fee:€1.58m'
 '€1.20m' 'Loan fee:€700Th.' 'Loan fee:€300Th.' '€210Th.' '€25Th.'
 '€790Th.' '€450Th.' 'Loan fee:€260Th.' '€16.00m' '€7.00m' '€6.50m'
 '€2.25m' '€1.33m' '€945Th.' '€338Th.' '€115Th.' '€2.15m' '€1.85m'
 'Loan fee:€1.66m' '€1.16m' 'Loan fee:€1.10m' 'Loan fee:€800Th.' '€795Th.'
 '€750Th.' '€492Th.' '€479Th.' 'Loan fee:€330Th.' '€320Th.' '€289Th.'
 '€225Th.' '€93Th.' 'Loan fee:€50Th.' '€35Th.' '€29Th.' 'Loan fee:€15Th.'
 '€2.08m' '€1.55m' '€45Th.' '€11.82m' '€660Th.' '€261Th.'
 'Loan fee:€84Th.' '€663Th.' '€727Th.' '€900Th.' 'Loan fee:€99Th.'
 'Loan fee:€82Th.' '€2.20m' '€1.40m' '€473Th.' '€112Th.' 'Loan fee:€35Th.'
 '€23Th.' '€4.95m' '€2.80m' '€1.65m' '€1.44m' '€1.31m' '€168Th.' '€106Th.'
 '€1.64m' '€5.23m' '€2.31m' 'Loan fee:€900Th.' 'Loan fee:€26Th.'
 'Loan fee:€14Th.' 'Loan fee:€12Th.' '0' '€962Th.' '€291Th.'
 'Loan fee:€166Th.' '€830Th.' 'Loan fee:€58Th.' 'Loan fee:€10Th.' '€5.90m'
 '€2.73m' '€2.07m' '€315Th.' '€273Th.' '€3.64m' '€274Th.' '€403Th.'
 '€8Th.' '€4.90m' '€83Th.' '€318Th.' 'Loan fee:€180Th.' 'Loan fee:€638Th.'
 'Loan fee:€45Th.' '€2.59m' '€364Th.' '€140Th.' '€720Th.'
 'Loan fee:€36Th.' '€224Th.' '€182Th.' '€2.57m' '€114Th.' '€68Th.'
 '€530Th.' '€5Th.' 'Loan fee:€40Th.' 'Loan fee:€20Th.' '€2.11m'
```

```python
# first I will create a copy of the transfer_fee column to preserve the original data
transfers_updated['fee'] = transfers_updated['transfer_fee'].copy(deep=True)
transfers_updated['fee'] = transfers_updated['fee'].replace('draft', float(0.0))
transfers_updated['fee'] = transfers_updated['fee'].replace('loan transfer', float(0.0))
transfers_updated['fee'] = transfers_updated['fee'].replace('free transfer', float(0.0))
transfers_updated['fee'] = transfers_updated['fee'].replace('?', float(0.0))
transfers_updated['fee'] = transfers_updated['fee'].replace('-', float(0.0))
transfers_updated['fee'] = transfers_updated['fee'].replace(np.nan, float(0.0))

# next, we will remove the loan fee text
transfers_updated['fee'] = transfers_updated['fee'].replace({'Loan fee:': ''}, regex=True)

# we will also remove the Euro symbol
transfers_updated['fee'] = transfers_updated['fee'].replace({'€': ''}, regex=True)
```

```python
# now we will create a separate column for the units (millions or thousands)

# to define units of interest
unit = {"m":"millions","Th": "thousands"}

# create function to check if key is contained within string
def check_unit(x):
    for key in unit:
        if key.lower() in x.lower():
            return unit[key]
    return 'none'

# apply string to fee column
# if a fee has an m - column will indicate millions
# if a fee has Th - column will indicate thousands
transfers_updated.fee=transfers_updated.fee.astype(str)

transfers_updated['fee_unit'] = transfers_updated['fee'].map(lambda x: check_unit(x))
```

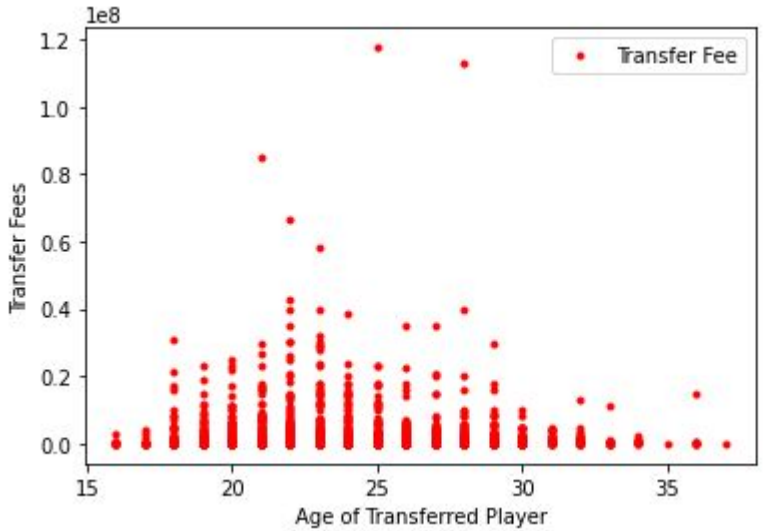
```python
# now we can update the fee column to remove the m and Th substrings from the column
# we will be left with the numerical value of the transfer
# then, we can use the unit column as a multiplyer for determining total value
transfers_updated['fee'] = transfers_updated['fee'].replace({'m': ''}, regex=True)
transfers_updated['fee'] = transfers_updated['fee'].replace({'Th.': ''}, regex=True)
transfers_updated.fee=transfers_updated.fee.astype(float)

def final_value(row):
    if row['fee_unit'] == 'millions':
        return row['fee']*1000000
    elif row['fee_unit'] == 'thousands':
        return row['fee']*1000
    else:
        return row['fee']
transfers_updated['transfer_fee_clean'] = transfers_updated.apply(lambda row: final_value(row), axis=1)
transfers_updated.transfer_fee_clean=transfers_updated.transfer_fee_clean.astype(float)
```

Age Analysis

- Players between 22 and 28 generally have higher values than those below or above in age
- Most transfers occurred for players aged 21-25
- Scatterplot helps show distribution of transfer fees by age which helps identify outliers

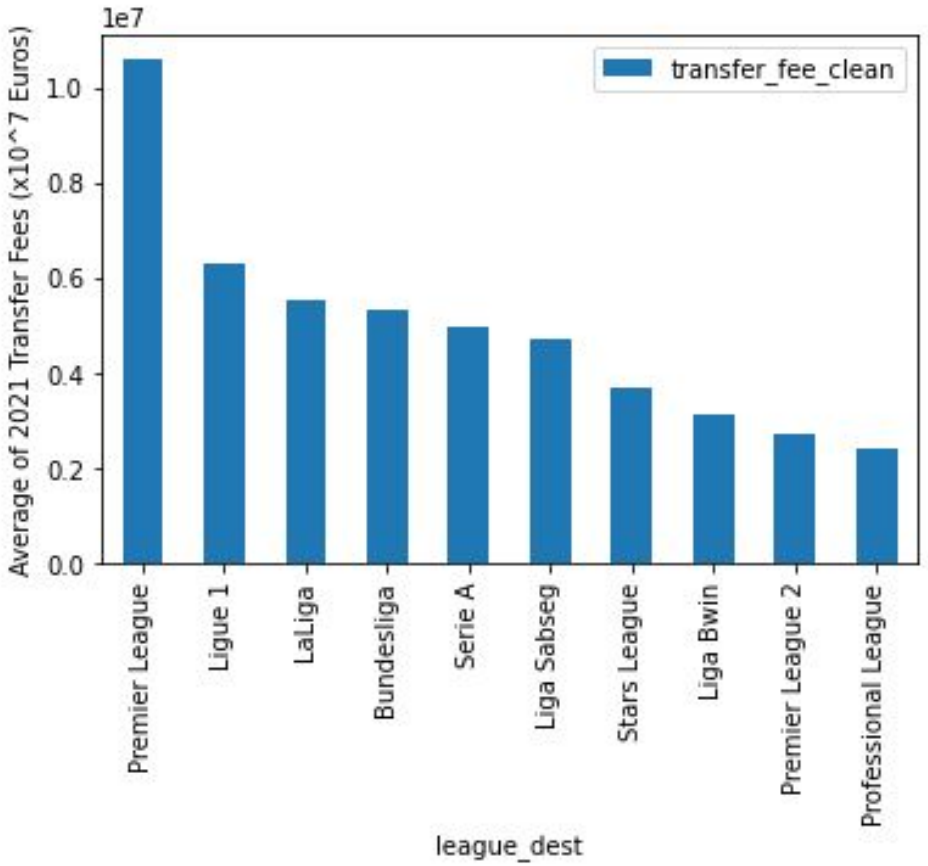
age_bins	transfer_fee_clean
(20, 25]	3159409.82
(10, 20]	2328610.22
(35, 45]	2294285.71
(25, 30]	2109305.41
(30, 35]	1249511.90

```
The following displays the number of
transfers that are not free transfers by age bins
(20, 25]    937
(25, 30]    573
(10, 20]    313
(30, 35]     84
(35, 45]      7
Name: age, dtype: int64
```

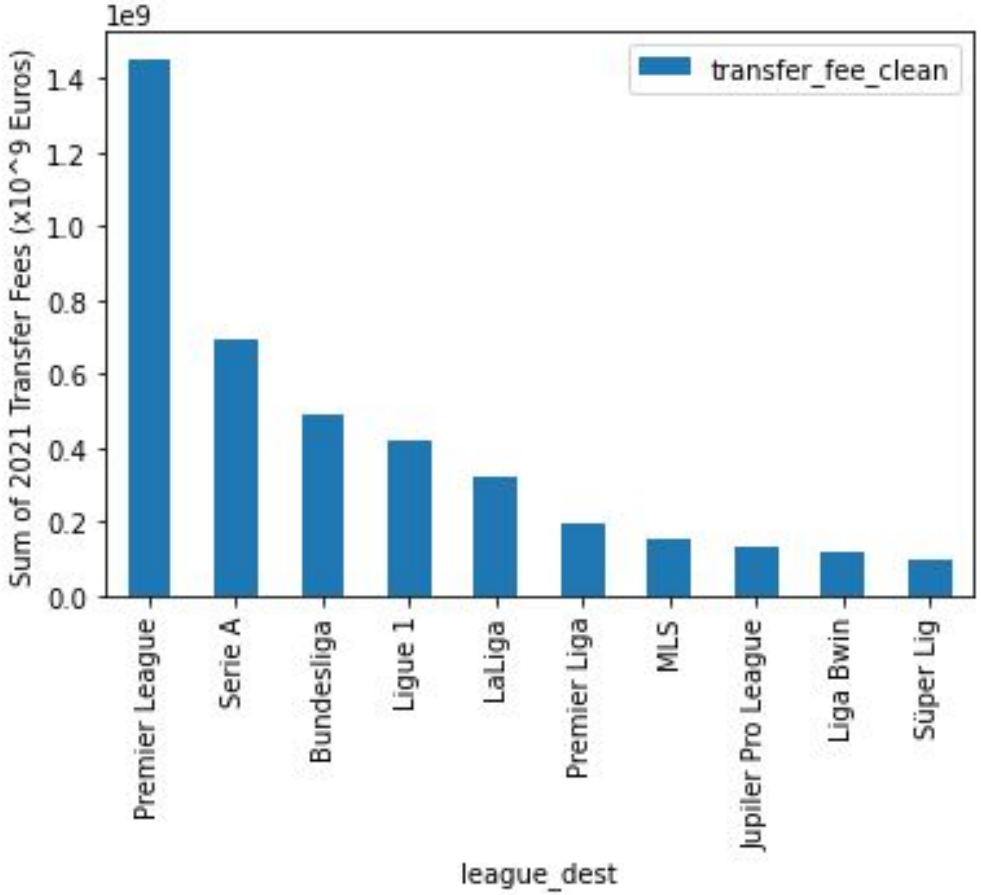


Destination League Analysis

Avg Transfer Value by League (Top 10)



Total 2021 Transfer Value by League (Top10)



Destination League Analysis

Avg Transfer Value by League (Top/Bottom 5)

league_dest	transfer_fee_clean
Premier League	10604744.53
Ligue 1	6320454.55
LaLiga	5526724.14
Bundesliga	5345347.83
Serie A	4951535.71
...	...
Cyprus	10000.00
Bangladesh PL	10000.00
Romania	5000.00
Serie D - H	2000.00
A Lyga	1000.00

133 rows × 1 columns

Max Transfer Value by League (Top 5)

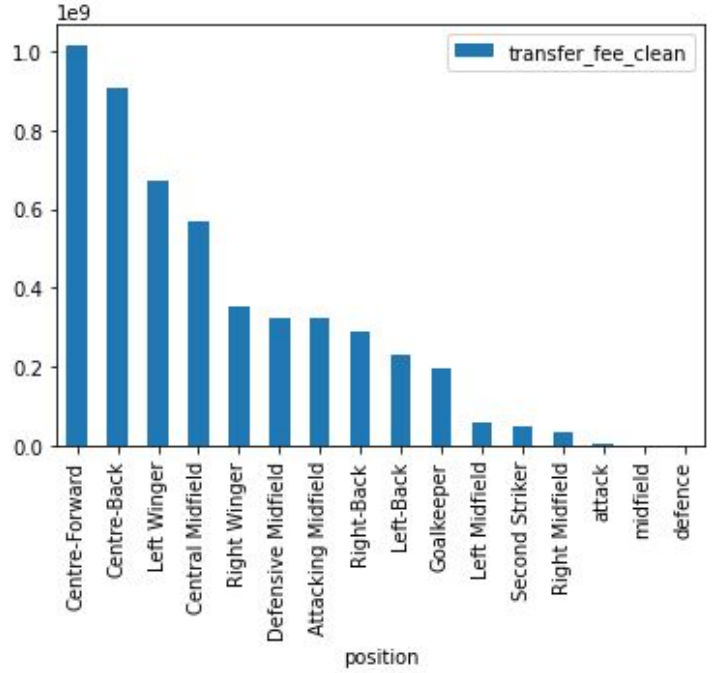
league_dest	transfer_fee_clean
Premier League	117500000.00
Ligue 1	66500000.00
Bundesliga	42500000.00
Serie A	40000000.00
LaLiga	35000000.00

Text goes here

Position Analysis

- 1,000,000,000 euros spent on center forwards!
- Center forwards had the highest total spend as a position group, but the highest average transfer fee was tied to left wingers

Total Spending by Position



Avg Transfer Value by Position