### **1. TV, halftime shows, and the Big Game**

*# Import pandas*

**import** **pandas** **as** **pd**

*# Load the CSV data into DataFrames*

super\_bowls = pd.read\_csv('datasets/super\_bowls.csv')

tv = pd.read\_csv('datasets/tv.csv')

halftime\_musicians = pd.read\_csv('datasets/halftime\_musicians.csv')

*# Display the first five rows of each DataFrame*

display(super\_bowls.head())

display(tv.head())

display(halftime\_musicians.head())

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **date** | **super\_bowl** | **venue** | **city** | **state** | **attendance** | **team\_winner** | **winning\_pts** | **qb\_winner\_1** | **qb\_winner\_2** | **coach\_winner** | **team\_loser** | **losing\_pts** | **qb\_loser\_1** | **qb\_loser\_2** | **coach\_loser** | **combined\_pts** | **difference\_pts** |
| **0** | 2018-02-04 | 52 | U.S. Bank Stadium | Minneapolis | Minnesota | 67612 | Philadelphia Eagles | 41 | Nick Foles | NaN | Doug Pederson | New England Patriots | 33 | Tom Brady | NaN | Bill Belichick | 74 | 8 |
| **1** | 2017-02-05 | 51 | NRG Stadium | Houston | Texas | 70807 | New England Patriots | 34 | Tom Brady | NaN | Bill Belichick | Atlanta Falcons | 28 | Matt Ryan | NaN | Dan Quinn | 62 | 6 |
| **2** | 2016-02-07 | 50 | Levi's Stadium | Santa Clara | California | 71088 | Denver Broncos | 24 | Peyton Manning | NaN | Gary Kubiak | Carolina Panthers | 10 | Cam Newton | NaN | Ron Rivera | 34 | 14 |
| **3** | 2015-02-01 | 49 | University of Phoenix Stadium | Glendale | Arizona | 70288 | New England Patriots | 28 | Tom Brady | NaN | Bill Belichick | Seattle Seahawks | 24 | Russell Wilson | NaN | Pete Carroll | 52 | 4 |
| **4** | 2014-02-02 | 48 | MetLife Stadium | East Rutherford | New Jersey | 82529 | Seattle Seahawks | 43 | Russell Wilson | NaN | Pete Carroll | Denver Broncos | 8 | Peyton Manning | NaN | John Fox | 51 | 35 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **super\_bowl** | **network** | **avg\_us\_viewers** | **total\_us\_viewers** | **rating\_household** | **share\_household** | **rating\_18\_49** | **share\_18\_49** | **ad\_cost** |
| **0** | 52 | NBC | 103390000 | NaN | 43.1 | 68 | 33.4 | 78.0 | 5000000 |
| **1** | 51 | Fox | 111319000 | 172000000.0 | 45.3 | 73 | 37.1 | 79.0 | 5000000 |
| **2** | 50 | CBS | 111864000 | 167000000.0 | 46.6 | 72 | 37.7 | 79.0 | 5000000 |
| **3** | 49 | NBC | 114442000 | 168000000.0 | 47.5 | 71 | 39.1 | 79.0 | 4500000 |
| **4** | 48 | Fox | 112191000 | 167000000.0 | 46.7 | 69 | 39.3 | 77.0 | 4000000 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | **super\_bowl** | **musician** | **num\_songs** |
| **0** | 52 | Justin Timberlake | 11.0 |
| **1** | 52 | University of Minnesota Marching Band | 1.0 |
| **2** | 51 | Lady Gaga | 7.0 |
| **3** | 50 | Coldplay | 6.0 |
| **4** | 50 | Beyoncé | 3.0 |

### **2. Taking note of dataset issues**

*# Summary of the TV data to inspect*

print(tv.describe())

print('**\n**')

*# Summary of the halftime musician data to inspect*

print(halftime\_musicians.describe())

### **3. Combined points distribution**

*# Import matplotlib and set plotting style*

**from** **matplotlib** **import** pyplot **as** plt

%matplotlib inline

plt.style.use('seaborn')

*# Plot a histogram of combined points*

plt.hist(super\_bowls.combined\_pts)

plt.xlabel('Combined Points')

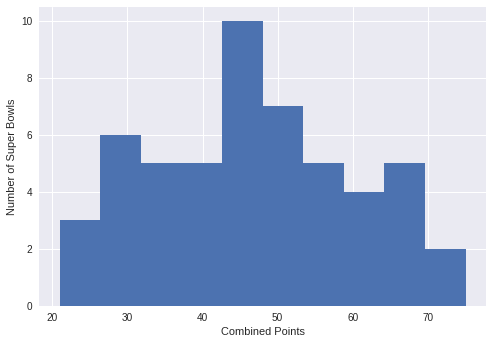
plt.ylabel('Number of Super Bowls')

plt.show()

*# Display the Super Bowls with the highest and lowest combined scores*

display(super\_bowls[super\_bowls['combined\_pts'] > 70])

display(super\_bowls[super\_bowls['combined\_pts'] < 25])



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **date** | **super\_bowl** | **venue** | **city** | **state** | **attendance** | **team\_winner** | **winning\_pts** | **qb\_winner\_1** | **qb\_winner\_2** | **coach\_winner** | **team\_loser** | **losing\_pts** | **qb\_loser\_1** | **qb\_loser\_2** | **coach\_loser** | **combined\_pts** | **difference\_pts** |
| **0** | 2018-02-04 | 52 | U.S. Bank Stadium | Minneapolis | Minnesota | 67612 | Philadelphia Eagles | 41 | Nick Foles | NaN | Doug Pederson | New England Patriots | 33 | Tom Brady | NaN | Bill Belichick | 74 | 8 |
| **23** | 1995-01-29 | 29 | Joe Robbie Stadium | Miami Gardens | Florida | 74107 | San Francisco 49ers | 49 | Steve Young | NaN | George Seifert | San Diego Chargers | 26 | Stan Humphreys | NaN | Bobby Ross | 75 | 23 |

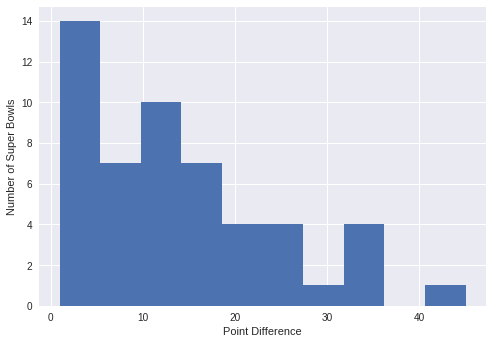
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **date** | **super\_bowl** | **venue** | **city** | **state** | **attendance** | **team\_winner** | **winning\_pts** | **qb\_winner\_1** | **qb\_winner\_2** | **coach\_winner** | **team\_loser** | **losing\_pts** | **qb\_loser\_1** | **qb\_loser\_2** | **coach\_loser** | **combined\_pts** | **difference\_pts** |
| **43** | 1975-01-12 | 9 | Tulane Stadium | New Orleans | Louisiana | 80997 | Pittsburgh Steelers | 16 | Terry Bradshaw | NaN | Chuck Noll | Minnesota Vikings | 6 | Fran Tarkenton | NaN | Bud Grant | 22 | 10 |
| **45** | 1973-01-14 | 7 | Memorial Coliseum | Los Angeles | California | 90182 | Miami Dolphins | 14 | Bob Griese | NaN | Don Shula | Washington Redskins | 7 | Bill Kilmer | NaN | George Allen | 21 | 7 |
| **49** | 1969-01-12 | 3 | Orange Bowl | Miami | Florida | 75389 | New York Jets | 16 | Joe Namath | NaN | Weeb Ewbank | Baltimore Colts | 7 | Earl Morrall | Johnny Unitas | Don Shula | 23 | 9 |

### **4. Point difference distribution**

*# Plot a histogram of point differences*

plt.hist(super\_bowls.difference\_pts)

plt.xlabel('Point Difference')

plt.ylabel('Number of Super Bowls')

plt.show()

*# Display the closest game(s) and biggest blowouts*

display(super\_bowls[super\_bowls['difference\_pts'] == 1])

display(super\_bowls[super\_bowls['difference\_pts'] >= 35])

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **date** | **super\_bowl** | **venue** | **city** | **state** | **attendance** | **team\_winner** | **winning\_pts** | **qb\_winner\_1** | **qb\_winner\_2** | **coach\_winner** | **team\_loser** | **losing\_pts** | **qb\_loser\_1** | **qb\_loser\_2** | **coach\_loser** | **combined\_pts** | **difference\_pts** |
| **27** | 1991-01-27 | 25 | Tampa Stadium | Tampa | Florida | 73813 | New York Giants | 20 | Jeff Hostetler | NaN | Bill Parcells | Buffalo Bills | 19 | Jim Kelly | NaN | Marv Levy | 39 | 1 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **date** | **super\_bowl** | **venue** | **city** | **state** | **attendance** | **team\_winner** | **winning\_pts** | **qb\_winner\_1** | **qb\_winner\_2** | **coach\_winner** | **team\_loser** | **losing\_pts** | **qb\_loser\_1** | **qb\_loser\_2** | **coach\_loser** | **combined\_pts** | **difference\_pts** |
| **4** | 2014-02-02 | 48 | MetLife Stadium | East Rutherford | New Jersey | 82529 | Seattle Seahawks | 43 | Russell Wilson | NaN | Pete Carroll | Denver Broncos | 8 | Peyton Manning | NaN | John Fox | 51 | 35 |
| **25** | 1993-01-31 | 27 | Rose Bowl | Pasadena | California | 98374 | Dallas Cowboys | 52 | Troy Aikman | NaN | Jimmy Johnson | Buffalo Bills | 17 | Jim Kelly | Frank Reich | Marv Levy | 69 | 35 |
| **28** | 1990-01-28 | 24 | Louisiana Superdome | New Orleans | Louisiana | 72919 | San Francisco 49ers | 55 | Joe Montana | NaN | George Seifert | Denver Broncos | 10 | John Elway | NaN | Dan Reeves | 65 | 45 |
| **32** | 1986-01-26 | 20 | Louisiana Superdome | New Orleans | Louisiana | 73818 | Chicago Bears | 46 | Jim McMahon | NaN | Mike Ditka | New England Patriots | 10 | Tony Eason | Steve Grogan | Raymond Berry | 56 | 36 |

### **5. Do blowouts translate to lost viewers?**

*# Join game and TV data, filtering out SB I because it was split over two networks*

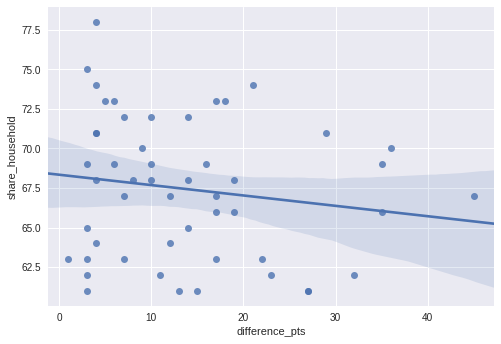
games\_tv = pd.merge(tv[tv['super\_bowl'] > 1], super\_bowls, on='super\_bowl')

*# Import seaborn*

**import** **seaborn** **as** **sns**

*# Create a scatter plot with a linear regression model fit*

sns.regplot(x='difference\_pts', y='share\_household', data=games\_tv)



### **6. Viewership and the ad industry over time**

*# Create a figure with 3x1 subplot and activate the top subplot*

plt.subplot(3, 1, 1)

plt.plot(tv.super\_bowl, tv.avg\_us\_viewers, color='#648FFF')

plt.title('Average Number of US Viewers')

*# Activate the middle subplot*

plt.subplot(3, 1, 2)

plt.plot(tv.super\_bowl, tv.rating\_household, color='#DC267F')

plt.title('Household Rating')

*# Activate the bottom subplot*

plt.subplot(3, 1, 3)

plt.plot(tv.super\_bowl, tv.rating\_household, color='#FFB000')

plt.title('Ad Cost')

plt.xlabel('SUPER BOWL')

*# Improve the spacing between subplots*

plt.tight\_layout()

### **7. Halftime shows weren’t always this great**

# Display all halftime musicians for Super Bowls up to and including Super Bowl XXVII

halftime\_musicians[halftime\_musicians.super\_bowl <= 27]

|  |  |  |  |
| --- | --- | --- | --- |
|  | **super\_bowl** | **musician** | **num\_songs** |
| **80** | 27 | Michael Jackson | 5.0 |
| **81** | 26 | Gloria Estefan | 2.0 |
| **82** | 26 | University of Minnesota Marching Band | NaN |
| **83** | 25 | New Kids on the Block | 2.0 |
| **84** | 24 | Pete Fountain | 1.0 |
| **85** | 24 | Doug Kershaw | 1.0 |
| **86** | 24 | Irma Thomas | 1.0 |
| **87** | 24 | Pride of Nicholls Marching Band | NaN |
| **88** | 24 | The Human Jukebox | NaN |
| **89** | 24 | Pride of Acadiana | NaN |
| **90** | 23 | Elvis Presto | 7.0 |
| **91** | 22 | Chubby Checker | 2.0 |
| **92** | 22 | San Diego State University Marching Aztecs | NaN |
| **93** | 22 | Spirit of Troy | NaN |
| **94** | 21 | Grambling State University Tiger Marching Band | 8.0 |
| **95** | 21 | Spirit of Troy | 8.0 |
| **96** | 20 | Up with People | NaN |
| **97** | 19 | Tops In Blue | NaN |
| **98** | 18 | The University of Florida Fightin' Gator March... | 7.0 |
| **99** | 18 | The Florida State University Marching Chiefs | 7.0 |
| **100** | 17 | Los Angeles Unified School District All City H... | NaN |
| **101** | 16 | Up with People | NaN |
| **102** | 15 | The Human Jukebox | NaN |
| **103** | 15 | Helen O'Connell | NaN |
| **104** | 14 | Up with People | NaN |
| **105** | 14 | Grambling State University Tiger Marching Band | NaN |
| **106** | 13 | Ken Hamilton | NaN |
| **107** | 13 | Gramacks | NaN |
| **108** | 12 | Tyler Junior College Apache Band | NaN |
| **109** | 12 | Pete Fountain | NaN |
| **110** | 12 | Al Hirt | NaN |
| **111** | 11 | Los Angeles Unified School District All City H... | NaN |
| **112** | 10 | Up with People | NaN |
| **113** | 9 | Mercer Ellington | NaN |
| **114** | 9 | Grambling State University Tiger Marching Band | NaN |
| **115** | 8 | University of Texas Longhorn Band | NaN |
| **116** | 8 | Judy Mallett | NaN |
| **117** | 7 | University of Michigan Marching Band | NaN |
| **118** | 7 | Woody Herman | NaN |
| **119** | 7 | Andy Williams | NaN |
| **120** | 6 | Ella Fitzgerald | NaN |
| **121** | 6 | Carol Channing | NaN |
| **122** | 6 | Al Hirt | NaN |
| **123** | 6 | United States Air Force Academy Cadet Chorale | NaN |
| **124** | 5 | Southeast Missouri State Marching Band | NaN |
| **125** | 4 | Marguerite Piazza | NaN |
| **126** | 4 | Doc Severinsen | NaN |
| **127** | 4 | Al Hirt | NaN |
| **128** | 4 | The Human Jukebox | NaN |
| **129** | 3 | Florida A&M University Marching 100 Band | NaN |
| **130** | 2 | Grambling State University Tiger Marching Band | NaN |
| **131** | 1 | University of Arizona Symphonic Marching Band | NaN |
| **132** | 1 | Grambling State University Tiger Marching Band | NaN |
| **133** | 1 | Al Hirt | NaN |

### **8. Who has the most halftime show appearances?**

*# Count halftime show appearances for each musician and sort them from most to least*

halftime\_appearances = halftime\_musicians.groupby('musician').count()['super\_bowl'].reset\_index()

halftime\_appearances = halftime\_appearances.sort\_values('super\_bowl', ascending=**False**)

*# Display musicians with more than one halftime show appearance*

halftime\_appearances[halftime\_appearances['super\_bowl'] > 1]

|  |  |  |
| --- | --- | --- |
|  | **musician** | **super\_bowl** |
| **28** | Grambling State University Tiger Marching Band | 6 |
| **104** | Up with People | 4 |
| **1** | Al Hirt | 4 |
| **83** | The Human Jukebox | 3 |
| **76** | Spirit of Troy | 2 |
| **25** | Florida A&M University Marching 100 Band | 2 |
| **26** | Gloria Estefan | 2 |
| **102** | University of Minnesota Marching Band | 2 |
| **10** | Bruno Mars | 2 |
| **64** | Pete Fountain | 2 |
| **5** | Beyoncé | 2 |
| **36** | Justin Timberlake | 2 |
| **57** | Nelly | 2 |
| **44** | Los Angeles Unified School District All City H... | 2 |

### **9. Who performed the most songs in a halftime show?**

*# Filter out most marching bands*

no\_bands = halftime\_musicians[~halftime\_musicians.musician.str.contains('Marching')]

no\_bands = no\_bands[~no\_bands.musician.str.contains('Spirit')]

*# Plot a histogram of number of songs per performance*

most\_songs = int(max(no\_bands['num\_songs'].values))

plt.hist(no\_bands.num\_songs.dropna(), bins=most\_songs)

plt.xlabel('Number of Songs Per Halftime Show Performance')

plt.ylabel('Number of Musicians')

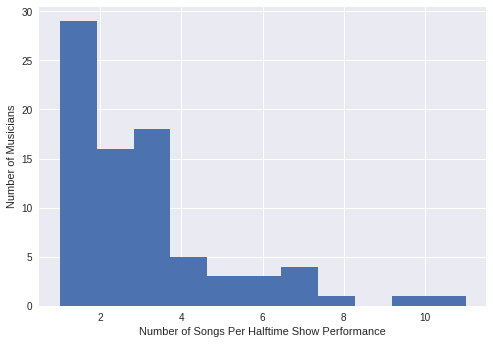
plt.show()

*# Sort the non-band musicians by number of songs per appearance...*

no\_bands = no\_bands.sort\_values('num\_songs', ascending=**False**)

*# ...and display the top 15*

display(no\_bands.head(15))



|  |  |  |  |
| --- | --- | --- | --- |
|  | **super\_bowl** | **musician** | **num\_songs** |
| **0** | 52 | Justin Timberlake | 11.0 |
| **70** | 30 | Diana Ross | 10.0 |
| **10** | 49 | Katy Perry | 8.0 |
| **2** | 51 | Lady Gaga | 7.0 |
| **90** | 23 | Elvis Presto | 7.0 |
| **33** | 41 | Prince | 7.0 |
| **16** | 47 | Beyoncé | 7.0 |
| **14** | 48 | Bruno Mars | 6.0 |
| **3** | 50 | Coldplay | 6.0 |
| **25** | 45 | The Black Eyed Peas | 6.0 |
| **20** | 46 | Madonna | 5.0 |
| **30** | 44 | The Who | 5.0 |
| **80** | 27 | Michael Jackson | 5.0 |
| **64** | 32 | The Temptations | 4.0 |
| **36** | 39 | Paul McCartney | 4.0 |

### **10. Conclusion**

*# 2018-2019 conference champions*

patriots = 'New England Patriots'

rams = 'Los Angeles Rams'

*# Who will win Super Bowl LII?*

super\_bowl\_LII\_winner = rams

print('The winner of Super Bowl LII will be the', super\_bowl\_LII\_winner)

The winner of Super Bowl LII will be the Los Angeles Rams