Data Analysis Problem on Data Wrangling and Exploring 1. Import the necessary libraries import pandas as pd In [2]: 2. Import the dataset from this address. import pandas as pd olympic_Ddata=pd.read_csv("https://raw.githubusercontent.com/svkarthik86/Advance-Python-Numpy/main/olympics_dirty.csv",skiprows=4) olympic_Ddata Available_City Edition_year Sport_name Discipline Athlete NOC Gender Medal Out[1]: Event Event_gender 0 Athens 1896 Aquatics Swimming HAJOS, Alfred HUN Men 100m freestyle M Gold Athens 1896 HERSCHMANN, Otto AUT Silver 1 Aquatics Swimming Men 100m freestyle M 2 DRIVAS, Dimitrios Athens 1896 Aquatics Swimming GRE Men 100m freestyle for sailors Bronze 3 Athens 1896 Aquatics Swimming MALOKINIS, Ioannis GRE 100m freestyle for sailors M NaN Men 4 Swimming Athens 1896 Aquatics CHASAPIS, Spiridon NaN 100m freestyle for sailors Μ Silver 29211 2008 Wrestling Wrestling Gre-R ENGLICH, Mirko 84 - 96kg Silver Beijing GER Men M 29212 Beijing 2008 Wrestling Gre-R MIZGAITIS, Mindaugas LTU 96 - 120kg Wrestling Men M Bronze 29213 2008 96 - 120kg Beijing Wrestling Wrestling Gre-R PATRIKEEV, Yuri ARM Men M Bronze 29214 Beijing 2008 Wrestling Gre-R LOPEZ, Mijain CUB 96 - 120kg Gold Wrestling Men M 29215 Beijing 2008 Wrestling Wrestling Gre-R BAROEV, Khasan RUS Men 96 - 120kg Μ Silver 29216 rows × 10 columns 3. Assign it to a variable called odata. In [3]: odata=olympic_Ddata odata Available_City Edition_year Sport_name Out[3]: **Discipline** Athlete NOC Gender Event Event_gender Medal 0 Athens 1896 Aquatics Swimming HAJOS, Alfred HUN Men 100m freestyle Gold Athens 1 1896 Swimming HERSCHMANN, Otto 100m freestyle Silver Aquatics Men M 2 Athens 1896 Aquatics Swimming DRIVAS, Dimitrios GRE 100m freestyle for sailors M Bronze 3 **Athens** 1896 Aquatics Swimming MALOKINIS, Ioannis GRE Men 100m freestyle for sailors M NaN 4 1896 Athens Aquatics Swimming CHASAPIS, Spiridon NaN 100m freestyle for sailors M Silver 29211 Beijing 2008 Wrestling Wrestling Gre-R ENGLICH, Mirko GER Men 84 - 96kg Μ Silver Beijing 2008 29212 Wrestling Gre-R MIZGAITIS, Mindaugas LTU Men 96 - 120kg M Bronze Wrestling 29213 Beijing 2008 Wrestling Wrestling Gre-R PATRIKEEV, Yuri ARM Men 96 - 120kg M Bronze 29214 2008 Wrestling Gre-R LOPEZ, Mijain CUB 96 - 120kg Gold Beijing Wrestling Men 29215 2008 96 - 120kg Beijing Wrestling Wrestling Gre-R BAROEV, Khasan RUS Men M Silver 29216 rows × 10 columns 4. show the first 5 entries In [5]: odata.head(5) Out[5]: Available_City Edition_year Sport_name Discipline Athlete NOC Gender Event Event_gender Medal 100m freestyle 1896 HAJOS, Alfred Gold Athens Aquatics Swimming HUN Men Athens 1896 Aquatics Swimming HERSCHMANN, Otto Men 100m freestyle Silver 2 1896 Aquatics Swimming Athens DRIVAS, Dimitrios **GRE** Men 100m freestyle for sailors M Bronze 100m freestyle for sailors Athens 1896 Aquatics Swimming MALOKINIS, Ioannis GRE Men NaN 4 Athens 1896 CHASAPIS, Spiridon 100m freestyle for sailors M Silver Aquatics Swimming NaN Men 5. What is the number of observations in the dataset? In [6]: odata.size 292160 Out[6]: 6. What is the number of columns in the dataset? odata.shape In [16]: (29216, 10)Out[16]: 7. Print the name of all the columns. odata.columns In [7]: Index(['Available_City', 'Edition_year', 'Sport_name', 'Discipline', 'Athlete', Out[7]: 'NOC', 'Gender', 'Event', 'Event_gender', 'Medal'], dtype='object') 8. How is the dataset indexed? In [18]: type(odata) pandas.core.frame.DataFrame Out[18]: In [19] odata.index RangeIndex(start=0, stop=29216, step=1) Out[19] 9. Rename the column as Available_City to City, Edition_year to Edition, Sport_name to Sport import pandas as pd olympic_Ddata=pd.read_csv("https://raw.githubusercontent.com/svkarthik86/Advance-Python-Numpy/main/olympics_dirty.csv",skiprows=4) olympic_Ddata Available_City Edition_year Sport_name Discipline Athlete NOC Gender Event Event_gender Medal Out[39]: 0 1896 Swimming HAJOS, Alfred HUN Gold Athens Aquatics Men 100m freestyle M 1 Athens 1896 Swimming HERSCHMANN, Otto AUT 100m freestyle M Silver Aquatics Men 2 1896 DRIVAS, Dimitrios GRE 100m freestyle for sailors Athens Aquatics Swimming Men Bronze 3 1896 Athens Aquatics Swimming MALOKINIS, Ioannis GRE 100m freestyle for sailors NaN Men 4 1896 CHASAPIS, Spiridon 100m freestyle for sailors Athens Aquatics Swimming NaN Men M Silver Beijing 29211 2008 Wrestling Wrestling Gre-R ENGLICH, Mirko GER Men 84 - 96kg M Silver 29212 2008 Beijing Wrestling Wrestling Gre-R MIZGAITIS, Mindaugas LTU 96 - 120kg Men M Bronze 29213 2008 PATRIKEEV, Yuri ARM 96 - 120kg Beijing Wrestling Wrestling Gre-R Men M Bronze 29214 2008 Wrestling Gold Beijing Wrestling Gre-R LOPEZ, Mijain CUB 96 - 120kg M Men 29215 Beijing 2008 Wrestling Wrestling Gre-R BAROEV, Khasan RUS 96 - 120kg Silver Μ 29216 rows × 10 columns odata=olympic_Ddata odata.rename(columns={' City':'Available_City',' Edition':'Edition_year',' Sport':'Sport_name'},inplace=True) odata Athlete NOC Gender Available_City Edition_year Sport_name Out[41]: Discipline Event Event_gender Medal 0 Athens 1896 Swimming HAJOS, Alfred HUN Men 100m freestyle Gold Aquatics Μ 1 Athens 1896 Aquatics Swimming HERSCHMANN, Otto AUT Men 100m freestyle Silver M Bronze 2 Athens 1896 Aquatics Swimming DRIVAS, Dimitrios GRE Men 100m freestyle for sailors 3 Athens 1896 Aquatics Swimming MALOKINIS, Ioannis GRE 100m freestyle for sailors NaN Swimming Men 100m freestyle for sailors 4 Athens 1896 Aquatics CHASAPIS, Spiridon NaN Μ Silver 29211 Beijing 2008 Wrestling Wrestling Gre-R ENGLICH, Mirko GER Men 84 - 96kg Μ Silver 96 - 120kg 29212 Beijing 2008 Wrestling Wrestling Gre-R MIZGAITIS, Mindaugas LTU Men M Bronze 29213 2008 Wrestling Wrestling Gre-R PATRIKEEV, Yuri ARM 96 - 120kg Bronze Beijing Men M 29214 Beijing 2008 Wrestling Wrestling Gre-R LOPEZ, Mijain CUB Men 96 - 120kg Gold 29215 Beijing 2008 Wrestling Wrestling Gre-R BAROEV, Khasan RUS Men 96 - 120kg Μ Silver 29216 rows × 10 columns 10. How many Athletes own the Gold medal? odata[(odata.Medal=='Gold')]["Athlete"].value_counts() In [35]: PHELPS, Michael Out[35]: LATYNINA, Larisa 9 9 SPITZ, Mark NURMI, Paavo 9 LEWIS, Carl LISENKO-SHEVTSOVA, Lyudmila 1 THOMPSON, Donald James 1 HALBERG, Murray Gordon 1 YOUNG, Earl Verdelle 1 LOPEZ, Mijain 1 Name: Athlete, Length: 7840, dtype: int64 11. Display the male gold medal winners for the 100m Track? In [42]: od1=odata[(odata.Medal=='Gold')&(odata.Gender=='Men')&(odata.Event=='100m')&(odata.Discipline=='Athletics')] od1 Out[42]: Available_City Edition_year Sport_name Discipline Athlete NOC Gender Event Event_gender Medal 13 1896 USA Gold Athens Athletics Athletics BURKE, Thomas Men 100m M 231 **Paris** 1900 **Athletics** USA 100m Gold Athletics JARVIS, Francis Men 737 1904 **Athletics** USA 100m Gold St Louis Athletics HAHN, Archie Men M 1191 London 1908 Athletics **Athletics** WALKER, Reginald RSA Men 100m Gold 2022 1912 Gold Stockholm Athletics Athletics CRAIG, Ralph USA 100m M Men 2996 1920 Athletics Athletics PADDOCK, Charles USA Gold Antwerp Men 100m 4236 1924 **Athletics** GBR M Gold Paris Athletics ABRAHAMS, Harold Men 100m 5095 Amsterdam 1928 WILLIAMS, Percy 100m Gold Athletics Athletics CAN Men 5806 Los Angeles 1932 Athletics Athletics TOLAN, Eddie USA Men 100m M Gold 6427 Berlin 1936 OWENS, Jesse USA 100m Gold Athletics **Athletics** Men 7302 1948 DILLARD, Harrison Men London Athletics **Athletics** USA 100m M Gold 8121 Helsinki 1952 **Athletics** REMIGINO, Lindy USA 100m Μ Gold Athletics Men Athletics MORROW, Robert Joseph 9009 1956 Gold Melbourne / Stockholm Athletics USA Men 100m Μ Athletics 9924 Rome 1960 Athletics HARY, Armin EUA Men 100m Gold 10823 1964 Athletics HAYES, Robert USA 100m M Gold Tokyo Athletics Men 11865 Mexico 1968 Athletics HINES, James Ray USA 100m Gold Athletics Men 12902 1972 Athletics **Athletics** BORZOV, Valery **URS** 100m M Gold Munich Men 14069 1976 CRAWFORD, Hasely Montreal **Athletics Athletics** TRI Men 100m M Gold Gold 15374 1980 Athletics **Athletics** WELLS, Allan GBR 100m M Moscow Men 16794 1984 LEWIS, Carl USA Gold Los Angeles **Athletics Athletics** Men 100m M 18284 1988 LEWIS, Carl USA 100m M Gold Seoul Athletics Athletics Men 1992 19859 Barcelona Athletics **Athletics** CHRISTIE, Linford GBR Men 100m Μ Gold 21598 Gold Atlanta 1996 Athletics **Athletics** BAILEY, Donovan CAN Men 100m M 23521 Sydney 2000 Athletics **Athletics** GREENE, Maurice USA Men 100m Gold 25539 2004 Athletics USA 100m Μ Gold Athens Athletics GATLIN, Justin Men 27552 2008 Beijing Athletics **Athletics** BOLT, Usain JAM Men 100m M Gold 1. Which three countries have won the most medals in recent years (from 1984 to 2008)? odata[(odata.Edition_year>=1984)&(odata.Edition_year<=2008)]["NOC"].value_counts().head(3) In [44] 1837 USA Out[44] **AUS** 762 691 **GER** Name: NOC, dtype: int64 13. Which country has won the most men's gold medals in singles badminton over the years? Sort the results alphabetically by the player's names. odata[(odata.Medal=='Gold')&(odata.Gender=='Men')&(odata.Event=='singles')&(odata.Discipline=='Badminton')].sort_values(['Athlete']) In [45]: Out[45]: Athlete NOC Gender Available_City Edition_year Sport_name Discipline Event Event_gender Medal 20045 Barcelona 1992 Badminton Badminton BUDI KUSUMA, Alan INA Men singles M Gold 25734 Athens 2004 Badminton Badminton HIDAYAT, Taufik INA singles Gold Men M Badminton Badminton HOYER-LARSEN, Poul Erik DEN 21787 Atlanta 1996 Gold Men singles M 23717 2000 Sydney Badminton Badminton JI, Xinpeng CHN Men singles M Gold 27741 Beijing 2008 Badminton Badminton LIN, Dan CHN Gold Men singles M 14. Display the women gold medal winners for the 100m Track? od1=odata[(odata.Medal=='Gold')&(odata.Gender=='Women')&(odata.Event=='100m')&(odata.Discipline=='Athletics')] In [46]: od1 Out[46]: Available_City Edition_year Sport_name Discipline Athlete NOC Gender Event Event_gender Medal ROBINSON, Elizabeth 5098 100m Gold 1928 Athletics USA Women W Amsterdam Athletics 5809 Los Angeles 1932 Athletics **Athletics** WALASIEWICZ, Stanislawa POL Women 100m W Gold STEPHENS, Helen Herring 6430 Berlin 1936 Athletics **Athletics** USA Women 100m W Gold **Athletics** 7305 London 1948 Athletics BLANKERS-KOEN, Fanny NED Women 100m W Gold 8124 Helsinki 1952 **Athletics** 100m Gold **Athletics** JACKSON, Marjorie AUS Women W Melbourne / Stockholm 9012 1956 Athletics Athletics **CUTHBERT**, Betty AUS Women 100m W Gold 9927 1960 100m Gold Rome **Athletics Athletics** RUDOLPH, Wilma USA Women W 10826 Tokyo 1964 **Athletics Athletics** TYUS, Wyomia USA Women 100m W Gold 11868 TYUS, Wyomia 100m Gold 1968 USA Women W Mexico Athletics Athletics STECHER, Renate GDR 12905 Munich 1972 Athletics Athletics Women 100m W Gold Athletics RICHTER-IRRGANG, Annegret 14072 1976 Athletics FRG 100m W Gold Montreal Women 15377 Moscow 1980 Athletics **Athletics** KONDRATIEVA, Liudmila URS Women 100m W Gold 16797 Los Angeles 1984 **Athletics** Athletics ASHFORD, Evelyn USA Women 100m W Gold 18287 1988 Athletics Athletics GRIFFITH-JOYNER, Florence USA Women 100m W Gold Seoul 19862 1992 Athletics Athletics DEVERS, Gail 100m W Gold Barcelona USA Women **Athletics** 21601 Atlanta 1996 **Athletics** DEVERS, Gail USA Women 100m W Gold Athens NESTERENKO, Yuliya 25542 2004 BLR 100m Gold Athletics **Athletics** Women W 27554 Beijing 2008 Athletics **Athletics** FRASER, Shelly-ann JAM Women 100m W Gold 15.list out silver medal winners at 2008 In [48]: odata.dropna() Edition_year Sport_name Athlete NOC Gender Event Event_gender Medal Out[48]: Available_City Discipline 0 Athens 1896 Aquatics Swimming HAJOS, Alfred HUN Men 100m freestyle Μ Gold 1896 Swimming Athens HERSCHMANN, Otto AUT Men 100m freestyle Μ Silver 1 Aquatics Swimming 2 Athens 1896 Aquatics DRIVAS, Dimitrios GRE Men 100m freestyle for sailors Bronze ANDREOU, Joannis GRE 7 1896 Athens Aquatics Men 1200m freestyle M Silver Swimming 9 Athens 1896 Aquatics Swimming NEUMANN, Paul Men 400m freestyle Μ Gold Wrestling Wrestling Gre-R 29211 Beijing 2008 ENGLICH, Mirko GER Men 84 - 96kg Μ Silver 29212 2008 96 - 120kg Beijing Wrestling Wrestling Gre-R MIZGAITIS, Mindaugas LTU Men M Bronze 29213 Beijing 2008 Wrestling Wrestling Gre-R PATRIKEEV, Yuri ARM Men 96 - 120kg Bronze 29214 2008 96 - 120kg Beijing Wrestling Wrestling Gre-R LOPEZ, Mijain CUB Men M Gold 29215 Beijing 2008 Wrestling Wrestling Gre-R BAROEV, Khasan RUS Men 96 - 120kg Silver 29205 rows × 10 columns In [51]: olympic_Ddata=pd.read_csv("https://raw.githubusercontent.com/svkarthik86/Advance-Python-Numpy/main/olympics_dirty.csv",skiprows=4) olympic_Ddata Available_City Edition_year Sport_name Discipline Athlete NOC Gender **Event_gender** Medal Out[51]: 0 Swimming HAJOS, Alfred HUN Athens 1896 Aquatics Men 100m freestyle Μ Gold 100m freestyle 1 Athens 1896 HERSCHMANN, Otto AUT Silver Aquatics Swimming Men 100m freestyle for sailors 2 1896 DRIVAS, Dimitrios GRE Athens Aquatics Swimming Men M Bronze 3 Athens 1896 Aquatics Swimming MALOKINIS, Ioannis GRE 100m freestyle for sailors NaN 4 1896 Swimming 100m freestyle for sailors Athens Aquatics CHASAPIS, Spiridon NaN Men Μ Silver 29211 2008 ENGLICH, Mirko 84 - 96kg Silver Beijing Wrestling Wrestling Gre-R GER Men Μ 29212 Wrestling Wrestling Gre-R MIZGAITIS, Mindaugas Beijing 2008 LTU Men 96 - 120kg M Bronze 29213 Beijing 2008 Wrestling Wrestling Gre-R PATRIKEEV, Yuri ARM Men 96 - 120kg M Bronze 29214 2008 Beijing Wrestling Wrestling Gre-R LOPEZ, Mijain CUB 96 - 120kg Gold Men 29215 2008 Wrestling Wrestling Gre-R BAROEV, Khasan RUS Beijing Men 96 - 120kg Μ Silver 29216 rows × 10 columns olympic_Ddata[(olympic_Ddata.Medal=='Silver')&(olympic_Ddata.Edition_year=='2008')].sort_values(['Athlete']) Available_City Edition_year Sport_name Discipline Athlete NOC Gender Event Event_gender Medal 16. Group by Countries and apply aggregate function min, max, count on over the year olympic_Ddata.groupby("NOC")["Edition_year"].agg([min, max,]) Out[59]: min max NOC **AFG** 2008 2008 **AHO** 1988 1988 ALG 1984 2008 **ANZ** 1908 1912 ARG 1924 2008 VIE 2000 2008 YUG 1924 2000 **ZAM** 1984 1996 **ZIM** 1980 2008 **ZZX** 1896 1904 138 rows × 2 columns