Documentation Data Analysis With Python

Project Specification: FIDE Chess Data Analytics

Project Overview

The goal of this project is to analyze the FIDE Chess Dataset using various data analytics techniques and tools. The dataset, stored in a CSV file named "fide_historical.csv," contains historical records of FIDE ratings for chess players.

Project Objectives

1. **Data Cleaning and Formatting:**
    - Import the necessary Python libraries: pandas, Seaborn, NumPy, matplotlib.pyplot, matplotlib.ticker, and datetime.
    - Read the "fide_historical.csv" file and load the data into a Pandas DataFrame.
    - Display the initial structure of the dataset in a tabular format.
2. **Data Preprocessing:**
    - Convert the 'ranking_date' column to datetime format and add a new column named 'date.'
3. **Individual Player Analysis:**
    - Display specific data for Magnus Carlsen:
        - In tabular format.
        - The record where Magnus Carlsen achieved his highest FIDE rating.

      - A line plot showing Magnus Carlsen’s FIDE rating over the years.
   - Display specific data for Wesley So and Kramnik Vladimir:
      - In tabular format.
      - Line plots showing FIDE ratings over the years.

4. **Comparison Plots:**
   - Plot FIDE ratings for Magnus Carlsen and Garry Kasparov in a single line plot.
   - Plot FIDE ratings for Magnus Carlsen and the rest of the world in a single line plot.
5. **Comparison of Multiple Players:**
   - Create an array of chess players who played after 2007-01-01 and do not include Magnus Carlsen.
   - Plot FIDE ratings for all selected players in a single line plot.

Deliverables

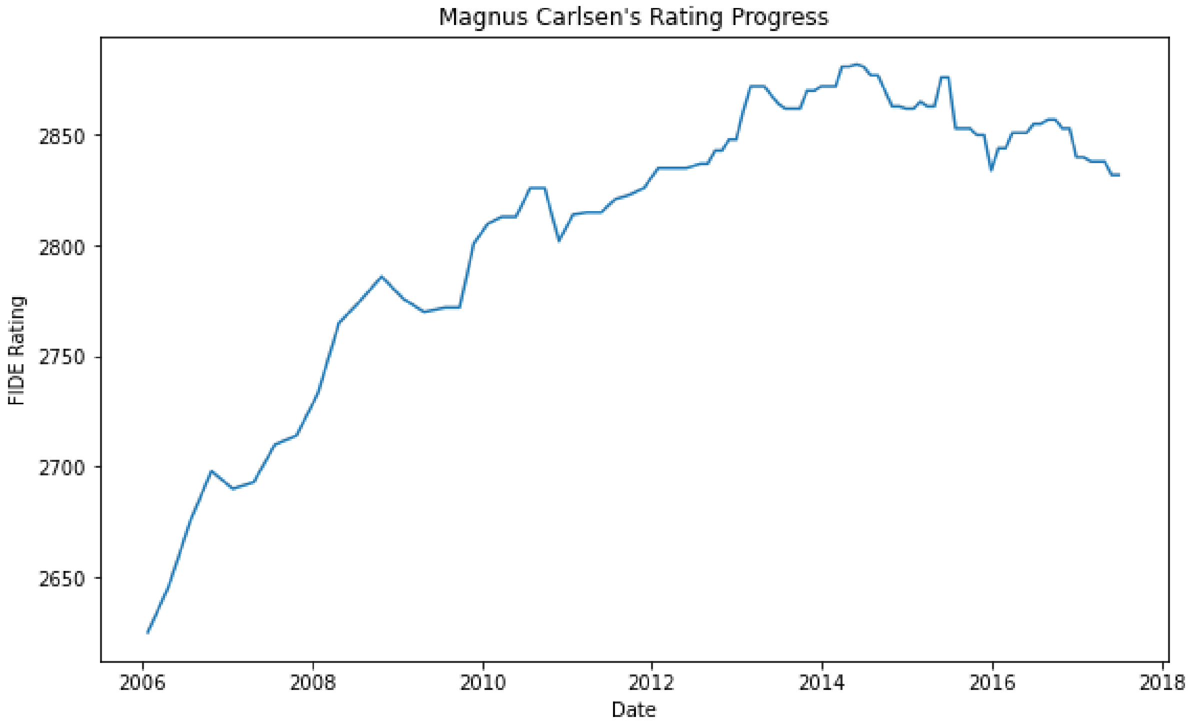
1. Jupyter Notebook containing the Python code for data analysis.
2. Visualizations for each analysis, including line plots, tables, and relevant figures.
3. A written report summarizing key findings, insights, and observations from the data.

Tools and Libraries

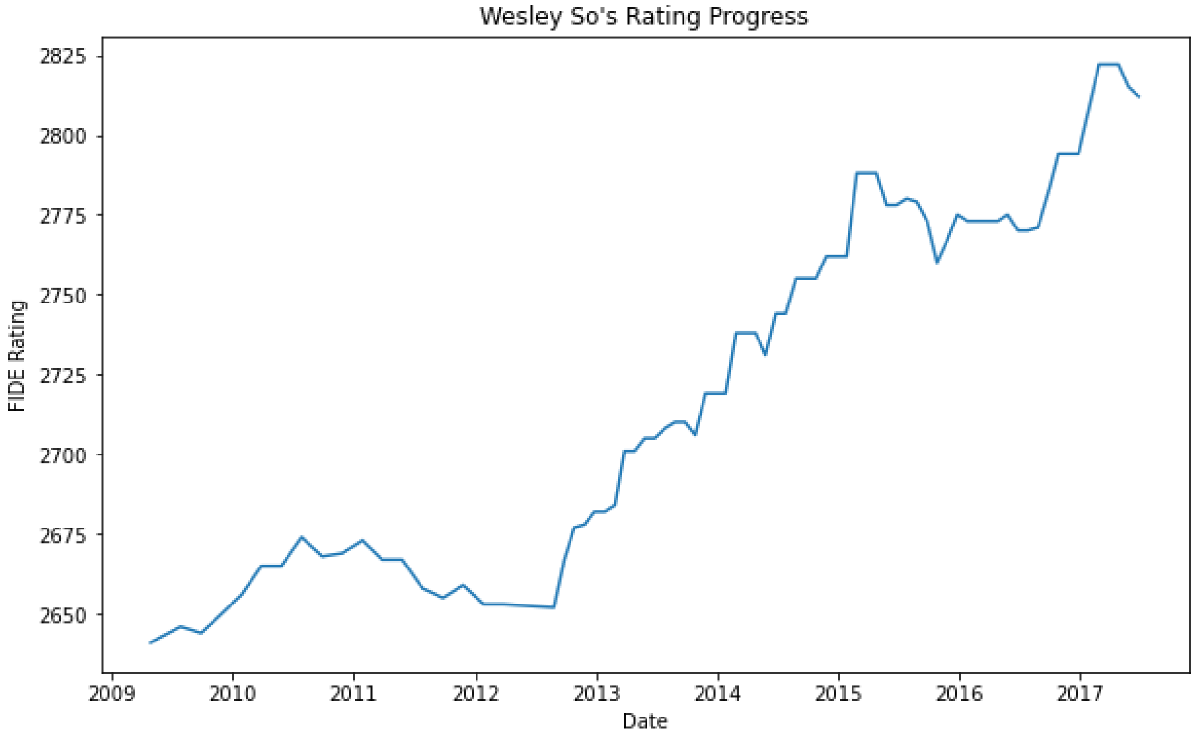
- Python 3.x
- Jupyter Notebook
- pandas
- Seaborn

- NumPy
- Matplotlib
- Spyder

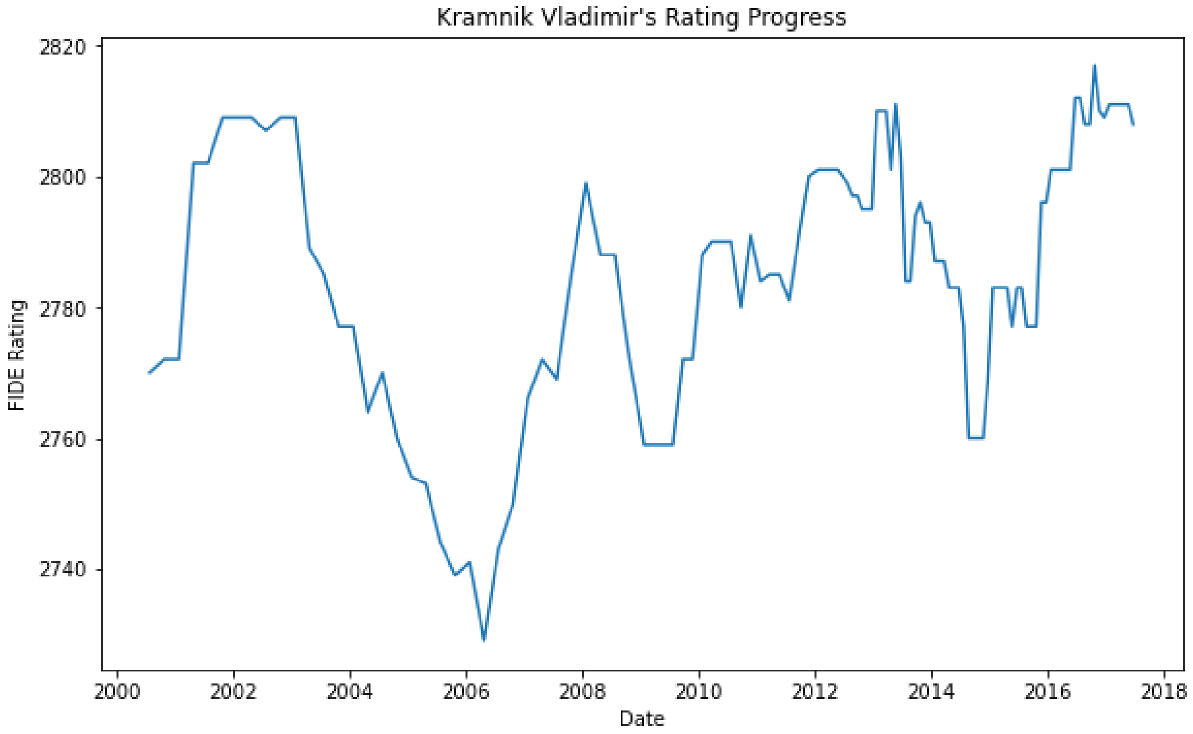
Question 1.4 c



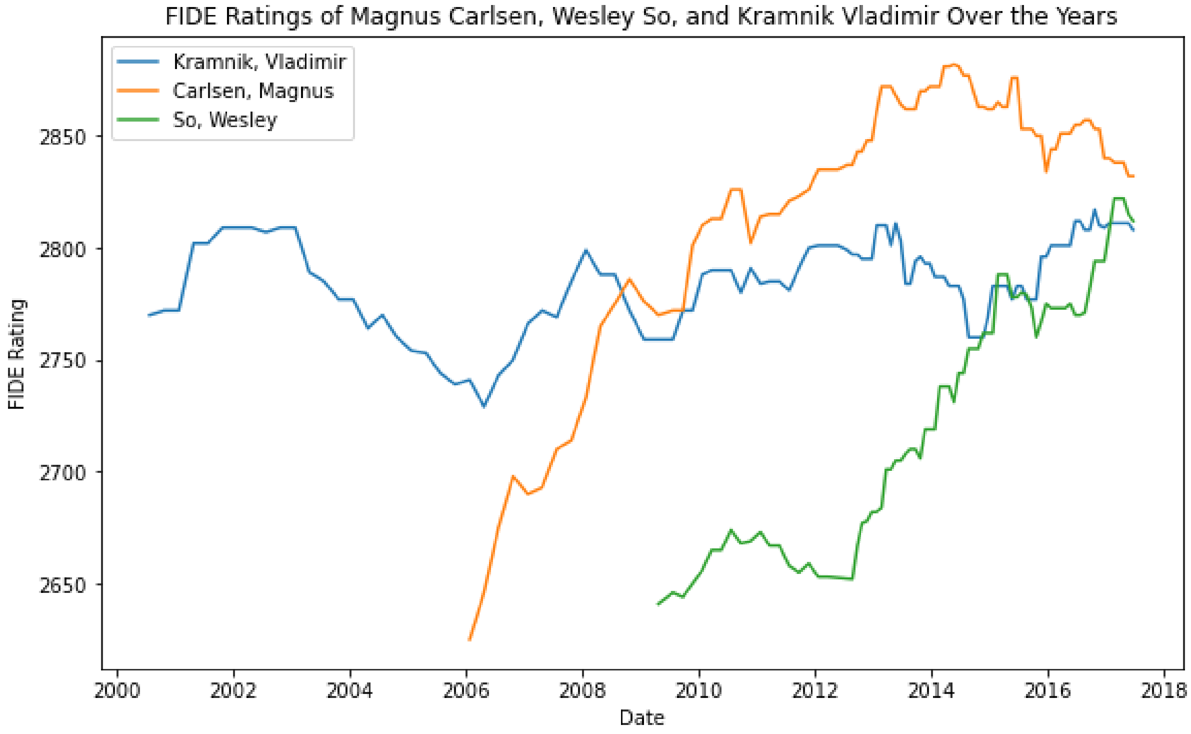
Question 1.4 e



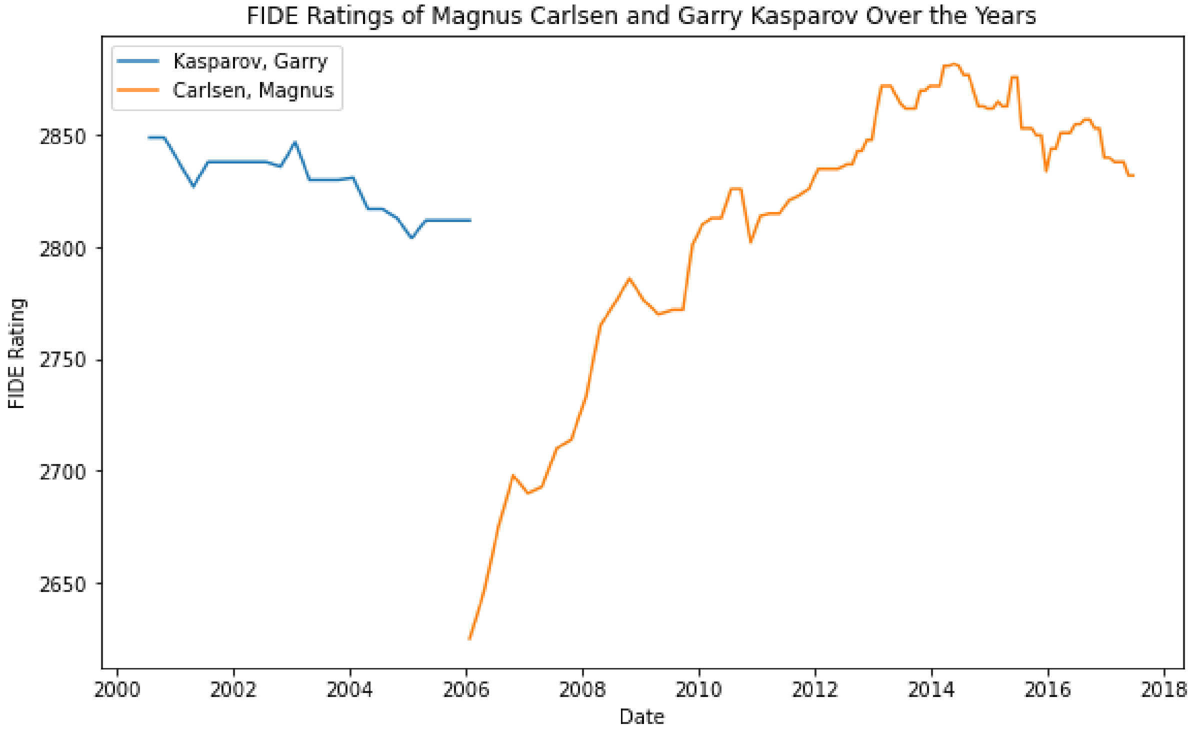
Question 1.4 f



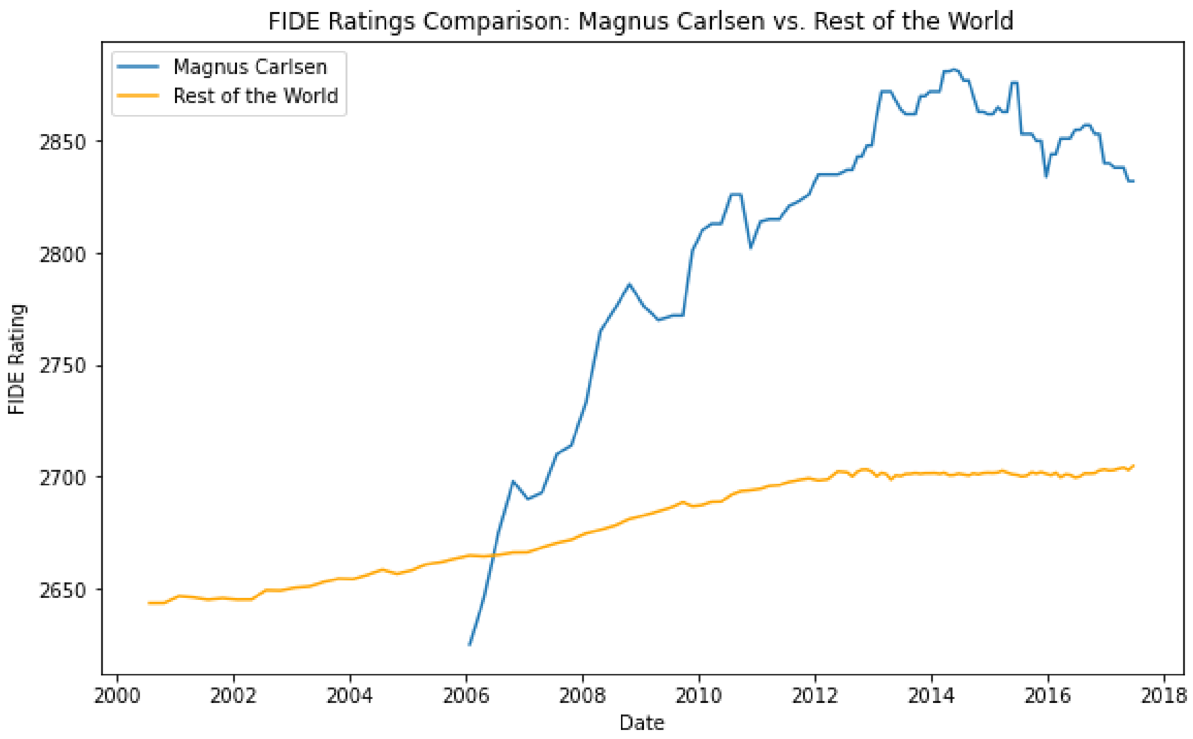
Question 1.4 g



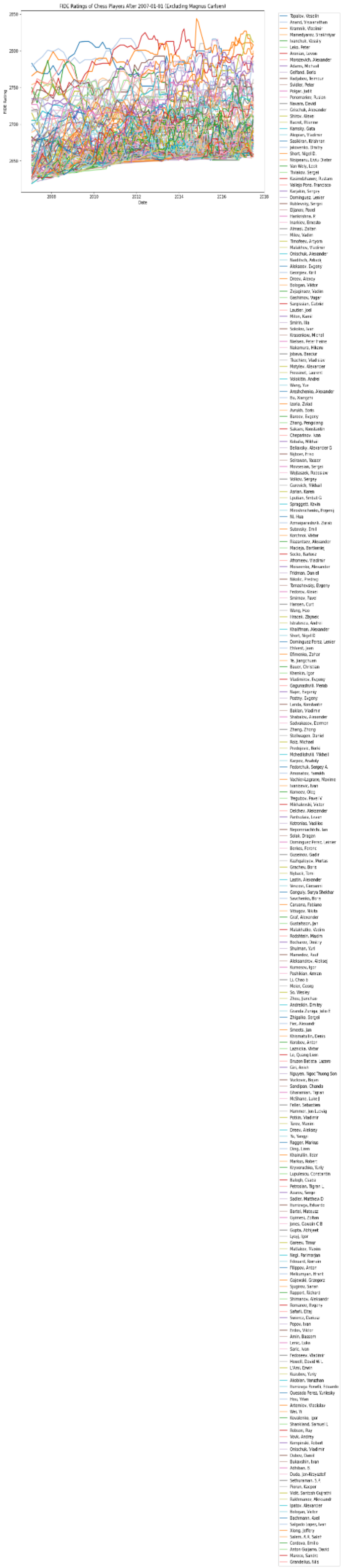
Question 1.5



Question 1.6


FIDE Ratings Comparison: Magnus Carlsen vs. Rest of the World
Magnus Carlsen
Rest of the World
FIDE Rating
2650
2700
2750
2800
2850
Date
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018

Question 1.8


FIDE Ratings of Chess Players After 2007-01-01 (Excluding Magnus Carlsen)
FIDE Rating
2850
2800
2750
2700
2650
Date
2008
2010
2012
2014
2016
2018
Topalov, Veselin
Anand, Viswanathan
Kramnik, Vladimir
Mamedyarov, Shakhriyar
Ivanchuk, Vassily
Leko, Peter
Aronian, Levon
Morozevich, Alexander
Adams, Michael
Gelfand, Boris
Radjabov, Teimour
Svidler, Peter
Polgar, Judit
Ponomariov, Ruslan
Navara, David
Grischuk, Alexander
Shirov, Alexei
Bacrot, Etienne
Kamsky, Gata
Akopian, Vladimir
Sasikiran, Krishnan
Jakovenko, Dmitry
Short, Nigel D.
Nisipeanu, Liviu-Dieter
Van Wely, Loek
Tiviakov, Sergei
Kasimdzhanov, Rustam
Vallejo Pons, Francisco
Karjakin, Sergey
Dominguez, Lenier
Rublevsky, Sergei
Eljanov, Pavel
Harikrishna, P.
Inarkiev, Ernesto
Almasi, Zoltan
Milov, Vadim
Timofeev, Artyom
Malakhov, Vladimir
Onischuk, Alexander
Naiditsch, Arkadij
Alekseev, Evgeny
Georgiev, Kiril
Dreev, Alexey
Bologan, Viktor
Zvjaginsev, Vadim
Gashimov, Vugar
Sargissian, Gabriel
Lautier, Joel
Miton, Kamil
Smirin, Ilia
Sokolov, Ivan
Krasenkow, Michal
Nielsen, Peter Heine
Nakamura, Hikaru
Jobava, Baadur
Tkachiev, Vladislav
Motylev, Alexander
Fressinet, Laurent
Volokitin, Andrei
Wang, Yue
Areshchenko, Alexander
Bu, Xiangzhi
Izoria, Zviad
Avrukh, Boris
Bareev, Evgeny
Zhang, Pengxiang
Sakaev, Konstantin
Cheparinov, Ivan
Kobalia, Mikhail
Beliavsky, Alexander G
Nijboer, Friso
Seirawan, Yasser
Movsesian, Sergei
Wojtaszek, Radoslaw
Volkov, Sergey
Gurevich, Mikhail
Asrian, Karen
Lputian, Smbat G
Spraggett, Kevin
Miroshnichenko, Evgenij
Ni, Hua
Azmaiparashvili, Zurab
Sutovsky, Emil
Korchnoi, Viktor
Riazantsev, Alexander
Macieja, Bartlomiej
Socko, Bartosz
Afromeev, Vladimir
Moiseenko, Alexander
Fridman, Daniel
Nikolic, Predrag
Tomashevsky, Evgeny
Fedorov, Alexei
Smirnov, Pavel
Hansen, Curt
Wang, Hao
Hracek, Zbynek
Istratescu, Andrei
Khalifman, Alexander
Short, Nigel D
Dominguez Perez, Lenier
Ehlvest, Jaan
Efimenko, Zahar
Ye, Jiangchuan
Bauer, Christian
Khenkin, Igor
Vladimirov, Evgeny
Gagunashvili, Merab
Najer, Evgeniy
Postny, Evgeny
Landa, Konstantin
Baklan, Vladimir
Shabalov, Alexander
Sadvakasov, Darmen
Zhang, Zhong
Stellwagen, Daniel
Roiz, Michael
Predojevic, Borki
Mchedlishvili, Mikheil
Karpov, Anatoly
Fedorchuk, Sergey A.
Amonatov, Farrukh
Vachier-Lagrave, Maxime
Ivanisevic, Ivan
Korneev, Oleg
Tregubov, Pavel V
Mikhalevski, Victor
Delchev, Aleksander
Pantsulaia, Levan
Kotronias, Vasilios
Nepomniachtchi, Ian
Solak, Dragan
Dominguez Perez, Leinier
Berkes, Ferenc
Guseinov, Gadir
Kazhgaleyev, Murtas
Grachev, Boris
Nyback, Tomi
Lastin, Alexander
Vescovi, Giovanni
Ganguly, Surya Shekhar
Savchenko, Boris
Caruana, Fabiano
Vitiugov, Nikita
Graf, Alexander
Gustafsson, Jan
Malakhatko, Vadim
Rodshtein, Maxim
Bocharov, Dmitry
Shulman, Yuri
Mamedov, Rauf
Aleksandrov, Aleksej
Kurnosov, Igor
Pashikian, Arman
Li, Chao b
Meier, Georg
So, Wesley
Zhou, Jianchao
Andreikin, Dmitry
Granda Zuniga, Julio E
Zhigalko, Sergei
Fier, Alexandr
Smeets, Jan
Khismatullin, Denis
Korobov, Anton
Laznicka, Viktor
Le, Quang Liem
Bruzon Batista, Lazaro
Giri, Anish
Nguyen, Ngoc Truong Son
Vuckovic, Bojan
Sandipan, Chanda
Gharamian, Tigran
McShane, Luke J
Feller, Sebastien
Hammer, Jon Ludvig
Potkin, Vladimir
Turov, Maxim
Dreev, Aleksey
Yu, Yangyi
Ragger, Markus
Ding, Liren
Khairullin, Ildar
Markus, Robert
Kryvoruchko, Yuriy
Lupulescu, Constantin
Balogh, Csaba
Petrosian, Tigran L
Azarov, Sergei
Sadler, Matthew D
Iturrizaga, Eduardo
Bartel, Mateusz
Gyimesi, Zoltan
Jones, Gawain C B
Gupta, Abhijeet
Lysyj, Igor
Gareev, Timur
Matlakov, Maxim
Negi, Parimarjan
Edouard, Romain
Filippov, Anton
Melkumyan, Hrant
Gajewski, Grzegorz
Sjugirov, Sanan
Rapport, Richard
Shimanov, Aleksandr
Romanov, Evgeny
Safarli, Eltaj
Swiercz, Dariusz
Popov, Ivan
Erdos, Viktor
Amin, Bassem
Lenic, Luka
Saric, Ivan
Fedoseev, Vladimir
Howell, David W L
L'Ami, Erwin
Kuzubov, Yuriy
Akobian, Varuzhan
Iturrizaga Bonelli, Eduardo
Quesada Perez, Yuniesky
Hou, Yifan
Artemiev, Vladislav
Wei, Yi
Kovalenko, Igor
Shankland, Samuel L
Robson, Ray
Vovk, Andrey
Kempinski, Robert
Onischuk, Vladimir
Dubov, Daniil
Bukavshin, Ivan
Adhiban, B.
Duda, Jan-Krzysztof
Sethuraman, S.P.
Piorun, Kacper
Vidit, Santosh Gujrathi
Rakhmanov, Aleksandr
Ipatov, Alexander
Bologan, Victor
Bachmann, Axel
Salgado Lopez, Ivan
Xiong, Jeffery
Salem, A.R. Saleh
Cordova, Emilio
Anton Guijarro, David
Mareco, Sandro
Grandelius, Nils

Source Code:

```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.ticker as ticker
from datetime import datetime

# Question 1.1
# Read the CSV file into a pandas DataFrame
file_path = "fide_historical.csv"
fide_data = pd.read_csv(file_path)

# Question 1.2
# Display the first few rows of the dataset to understand its structure
print(fide_data.head())

# Question 1.3
# Display the DataFrame in table format with a summary
print(fide_data)
print(fide_data.info())

# Convert the 'ranking_date' column to datetime format, handling errors
```

```
fide_data['date'] = pd.to_datetime(fide_data['ranking_date'], format='%d-%m-%y', errors='coerce')

# Display rows with NaT (parsing errors)
rows_with_errors = fide_data[fide_data['date'].isna()]
print("Rows with Parsing Errors:")
print(rows_with_errors)

# Drop rows with parsing errors or handle them based on your specific case
fide_data = fide_data.dropna(subset=['date'])

# Display the updated DataFrame with the new 'date' column
print("\nUpdated DataFrame:")
print(fide_data.head())


# Question 1.4
# a. Display only Magnus Carlsen's data in tabular format
magnus_data = fide_data[fide_data['name'] == 'Carlsen, Magnus']
print("Magnus Carlsen's Data:")
print(magnus_data)

# b. Display the record where Magnus Carlsen achieved his highest FIDE rating
```

```
max_rating_record = magnus_data.loc[magnus_data['rating'].idxmax()]
print("\nMagnus Carlsen's Record with Highest FIDE Rating:")
print(max_rating_record)

# c. Display Magnus Carlsen’s FIDE rating over the years using a lineplot
plt.figure(figsize=(10, 6))
sns.lineplot(x='date', y='rating', data=magnus_data)
plt.title("Magnus Carlsen's Rating Progress")
plt.xlabel("Date")
plt.ylabel("FIDE Rating")
plt.show()

# d. Display ratings in descending order from the date 2017-06-27
descending_ratings = fide_data[fide_data['date'] >= '2017-06-27'].sort_values(by=['date', 'rating'], ascending=[True, False])
print("\nRatings in Descending Order from 2017-06-27:")
print(descending_ratings)

# e. Display Wesley So’s FIDE rating over the years using a lineplot
wesley_data = fide_data[fide_data['name'] == 'So, Wesley']
plt.figure(figsize=(10, 6))
sns.lineplot(x='date', y='rating', data=wesley_data)
plt.title("Wesley So's Rating Progress")
plt.xlabel("Date")
plt.ylabel("FIDE Rating")
plt.show()

# f. Display Kramnik Vladimir’s FIDE rating over the years using a lineplot
kramnik_data = fide_data[fide_data['name'] == 'Kramnik, Vladimir']
plt.figure(figsize=(10, 6))
sns.lineplot(x='date', y='rating', data=kramnik_data)
plt.title("Kramnik Vladimir's Rating Progress")
plt.xlabel("Date")
plt.ylabel("FIDE Rating")
plt.show()

# g. Display all three chess players’ ratings in one lineplot
# Filter data for specific players
selected_players = ['Carlsen, Magnus', 'So, Wesley', 'Kramnik, Vladimir']
selected_data = fide_data[fide_data['name'].isin(selected_players)]

# Plot FIDE Ratings for the selected players
plt.figure(figsize=(10, 6))
sns.lineplot(x='date', y='rating', hue='name', data=selected_data)
plt.title("FIDE Ratings of Magnus Carlsen, Wesley So, and Kramnik Vladimir Over the Years")
plt.xlabel("Date")
plt.ylabel("FIDE Rating")
plt.legend()
plt.show()

# Question 1.5
# Filter data for Magnus Carlsen and Garry Kasparov
selected_players = ['Carlsen, Magnus', 'Kasparov, Garry']
selected_data = fide_data[fide_data['name'].isin(selected_players)]

# Plot FIDE Ratings for Magnus Carlsen and Garry Kasparov
plt.figure(figsize=(10, 6))
sns.lineplot(x='date', y='rating', hue='name', data=selected_data)
plt.title("FIDE Ratings of Magnus Carlsen and Garry Kasparov Over the Years")
plt.xlabel("Date")
plt.ylabel("FIDE Rating")
plt.legend()
plt.show()

 # Question 1.6
 # Filter data for Magnus Carlsen
carlsen_data = fide_data[fide_data['name'] == 'Carlsen, Magnus']
```

```
# Calculate the average or median rating for the rest of the world
world_data = fide_data[fide_data['name'] != 'Carlsen, Magnus']
world_average_rating = world_data.groupby('date')['rating'].mean()

# Plot FIDE Ratings for Magnus Carlsen and the rest of the world
plt.figure(figsize=(10, 6))
sns.lineplot(x='date', y='rating', data=carlsen_data, label='Magnus Carlsen')
sns.lineplot(x=world_average_rating.index, y=world_average_rating.values, label='Rest of the World', color='orange')
plt.title("FIDE Ratings Comparison: Magnus Carlsen vs. Rest of the World")
plt.xlabel("Date")
plt.ylabel("FIDE Rating")
plt.legend()
plt.show()

# Question 1.7
# Filter data for players after 2007-01-01 and not including Magnus Carlsen
filtered_players = fide_data[(fide_data['date'] > '2007-01-01') & (fide_data['name'] != 'Carlsen, Magnus')]

# Create an array of unique chess players meeting the criteria
```

```
unique_players_after_2007 = filtered_players['name'].unique()

# Display the array
print(unique_players_after_2007)

# Question 1.8
# Filter data for players after 2007-01-01 and not including Magnus Carlsen
filtered_players = fide_data[(fide_data['date'] > '2007-01-01') & (fide_data['name'] != 'Carlsen, Magnus')]

# Plot FIDE Ratings for all chess players meeting the criteria
plt.figure(figsize=(12, 8))
sns.lineplot(x='date', y='rating', hue='name', data=filtered_players, palette='tab20', linewidth=2)
plt.title("FIDE Ratings of Chess Players After 2007-01-01 (Excluding Magnus Carlsen)")
plt.xlabel("Date")
plt.ylabel("FIDE Rating")
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.show()
```