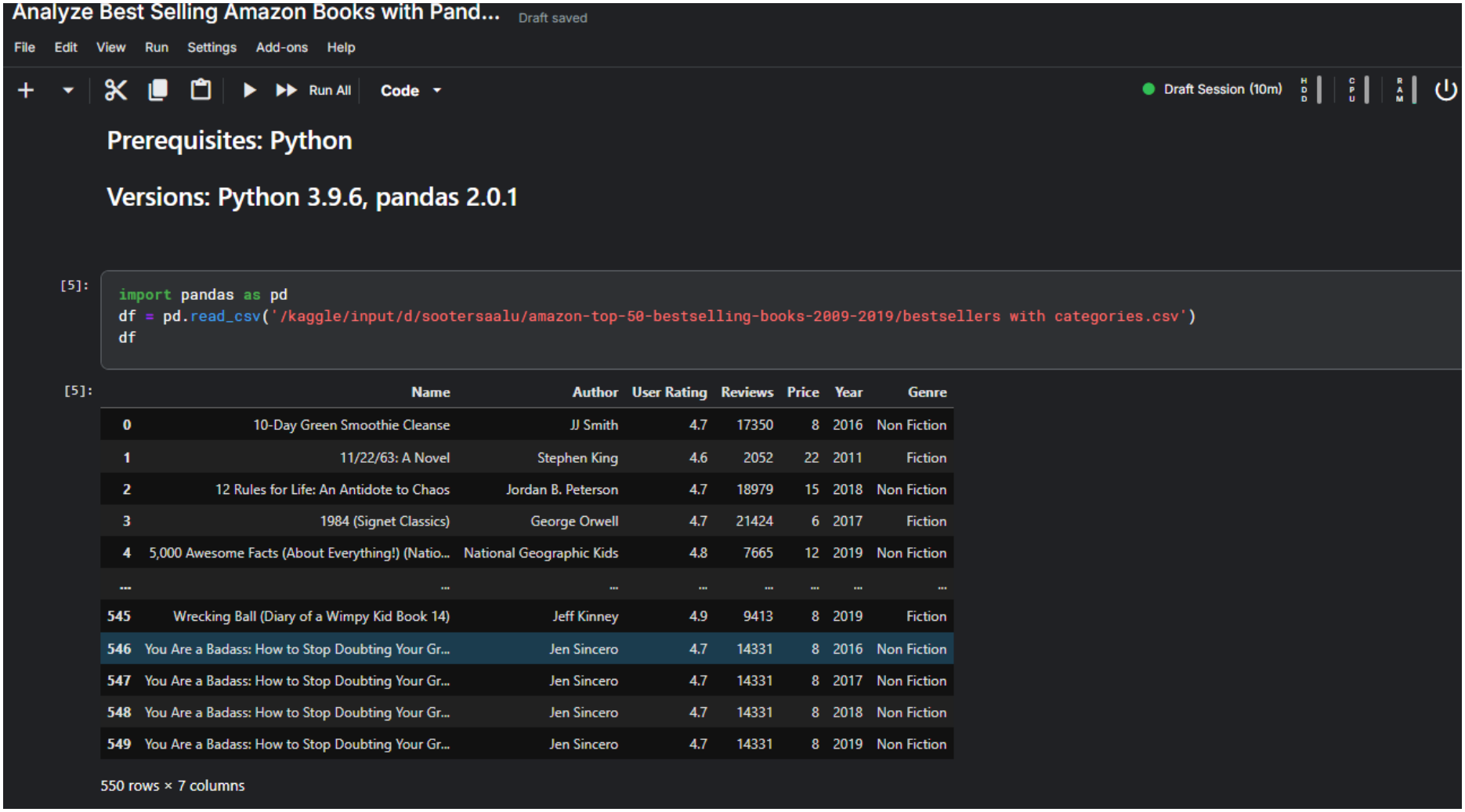
# Analyze Best Selling Amazon Books with Pandas




| | Name | Author | User Rating | Reviews | Price | Year | Genre |
|---|---|---|---|---|---|---|---|
| 0 | 10-Day Green Smoothie Cleanse | JJ Smith | 4.7 | 17350 | 8 | 2016 | Non Fiction |
| 1 | 11/22/63: A Novel | Stephen King | 4.6 | 2052 | 22 | 2011 | Fiction |
| 2 | 12 Rules for Life: An Antidote to Chaos | Jordan B. Peterson | 4.7 | 18979 | 15 | 2018 | Non Fiction |
| 3 | 1984 (Signet Classics) | George Orwell | 4.7 | 21424 | 6 | 2017 | Fiction |
| 4 | 5,000 Awesome Facts (About Everything!) (Natio... | National Geographic Kids | 4.8 | 7665 | 12 | 2019 | Non Fiction |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 545 | Wrecking Ball (Diary of a Wimpy Kid Book 14) | Jeff Kinney | 4.9 | 9413 | 8 | 2019 | Fiction |
| 546 | You Are a Badass: How to Stop Doubting Your Gr... | Jen Sincero | 4.7 | 14331 | 8 | 2016 | Non Fiction |
| 547 | You Are a Badass: How to Stop Doubting Your Gr... | Jen Sincero | 4.7 | 14331 | 8 | 2017 | Non Fiction |
| 548 | You Are a Badass: How to Stop Doubting Your Gr... | Jen Sincero | 4.7 | 14331 | 8 | 2018 | Non Fiction |
| 549 | You Are a Badass: How to Stop Doubting Your Gr... | Jen Sincero | 4.7 | 14331 | 8 | 2019 | Non Fiction |

550 rows × 7 columns

[6]:

```
print(df.head())

print(df.shape)

print(df.columns)

print(df.describe())
```

```
                                                Name  \
0                        10-Day Green Smoothie Cleanse
1                                    11/22/63: A Novel
2               12 Rules for Life: An Antidote to Chaos
3                                1984 (Signet Classics)
4  5,000 Awesome Facts (About Everything!) (Natio...

                     Author  User Rating  Reviews  Price  Year        Genre
0                 JJ Smith          4.7    17350      8  2016  Non Fiction
1              Stephen King          4.6     2052     22  2011      Fiction
2         Jordan B. Peterson          4.7    18979     15  2018  Non Fiction
3              George Orwell          4.7    21424      6  2017      Fiction
4  National Geographic Kids          4.8     7665     12  2019  Non Fiction
(550, 7)
Index(['Name', 'Author', 'User Rating', 'Reviews', 'Price', 'Year', 'Genre'], dtype='object')
       User Rating       Reviews       Price         Year
count   550.000000    550.000000  550.000000   550.000000
mean      4.618364  11953.281818   13.100000  2014.000000
std       0.226980  11731.132017   10.842262     3.165156
min       3.300000     37.000000    0.000000  2009.000000
25%       4.500000   4058.000000    7.000000  2011.000000
50%       4.700000   8580.000000   11.000000  2014.000000
75%       4.800000  17253.250000   16.000000  2017.000000
max       4.900000  87841.000000  105.000000  2019.000000
```

[9]:
```
df.drop_duplicates(inplace=True)
```

[10]:
```
df.rename(columns={"Name": "Title", "Year": "Publication Year", "User Rating": "Rating"}, inplace=True)
```

[11]:
```
df["Price"] = df["Price"].astype(float)
```

[12]:
```
author_counts = df['Author'].value_counts()
print(author_counts)
```

```
Author
Jeff Kinney                          12
Gary Chapman                         11
Rick Riordan                         11
Suzanne Collins                      11
American Psychological Association   10
                                     ..
Keith Richards                        1
Chris Cleave                          1
Alice Schertle                        1
Celeste Ng                            1
Adam Gasiewski                        1
Name: count, Length: 248, dtype: int64
```

[13]:
```
avg_rating_by_genre = df.groupby("Genre")["Rating"].mean()
print(avg_rating_by_genre)
```

```
Genre
Fiction        4.648333
Non Fiction    4.595161
Name: Rating, dtype: float64
```

[14]:
```
author_counts.head(10).to_csv("top_authors.csv")

avg_rating_by_genre.to_csv("avg_rating_by_genre.csv")
```