Python Panda Dataframes and SQL Equivalent

|  |  |  |  |
| --- | --- | --- | --- |
| Functionality | Python Pandas | SQL | Sample Output |
| Get the top rows (the first rows) from your source – such as the top five rows in your CSV. | data\_file\_pd.head()  *This defaults to the first 5 rows. Here is an example if you want to change the number of rows that it selects:*  data\_file\_pd.head(10) | SELECT TOP 5 \*  FROM data\_file\_pd  *For the top 10:*  SELECT TOP 10 \*  FROM data\_file\_pd |  |
| Refer a column name | Data\_file\_pd[“Amount”].head() | SELECT TOP 5 Amount  FROM data\_file\_pd |  |
| Show unique values | Data\_file\_pd[“Last Name”].unique | SELECT DISTINCT LastName  FROM data\_file\_pd |  |
| Count unique values |  | SELECT Gender, Count(data\_file\_pd.[Gender]) AS Counting  FROM data\_file\_pd  GROUP BY Gender  (or group by all columns) |  |
| Performing a calculation |  | SELECT TOP 5 Amount/1000 AS thousands\_of\_dollars  FROM data\_file\_pd |  |
| Rename a column |  | SELECT CharacterinShow AS Character, Color\_of\_hair AS Color, Height, Football\_Shaped\_Head AS FootballHead  FROM hey\_arnold |  |
| Organize columns |  |  |  |
| Select multiple columns |  |  |  |
|  |  | SELECT DISTINCT COUNT \*  (You do a count distinct) |  |
| Sum of multiple datasets (fields) | Goodreadssummary |  | See above |
| Locate (WHERE – filtering) |  | SELECT PhoneNumber FROM df WHERE LastName = Berry |  |
|  |  | SELECT id, first\_name, PhoneNumber  FROM df  WHERE Id IN (“Richardson”, “Berry”, “Hudson”, “McDonald”, “Morales”) |  |
|  |  | SELECT FirstName  FROM df.loc  WHERE firstname = “Billy” OR firstname = “Peter” |  |
|  |  | SELECT Film, IMDB, IMDB\_user\_vote\_count  FROM pd.loc  WHERE IMDB > 7 |  |
|  |  | How many crimes  SELECT COUNT (Offsense Type)  GROUP BY all other columns |  |
| Drop a column. |  |  |  |
|  |  |  |  |
| UPDATE | # Combining similar offenses together  no\_null\_crime\_df["Offense Type"] = no\_null\_crime\_df["Offense Type"].replace({"Commercial Sex Acts":"Prostitution","Assisting or Promoting Prostitution":"Prostitution"})  no\_null\_crime\_df |  |  |
|  |  | SELECT Country, State, SUM(duration(seconds))  FROM grouped\_international\_data  GROUP BY Country, State |  |
| Inner Join |  | SELECT \*  FROM info\_pd INNER JOIN items\_pd ON customer\_id = customer\_id |  |
|  |  | SELECT \*  FROM info\_pd LEFT OUTER JOIN items\_pd ON customer\_id = customer\_id |  |