sql-session-1

September 16, 2024

1 Import Libraries

```
[32]: import pandas as pd
      import sqlite3 as sql
```

2 Connecting to the Database

```
[33]: db = 'jobs.db'
```

```
[34]: conn = sql.connect(db)
```

```
[35]: conn
```

```
[35]: <sqlite3.Connection at 0x200c205c300>
```

3 Reading Query using Pandas

```
[10]: pd.read_sql_query('select * from sqlite_master', conn)
```

```
[10]:     type                    name      tbl_name  rootpage  \
      0  table            recent_grads  recent_grads         2
      1  index  ix_recent_grads_index  recent_grads         3

                                                       sql
      0  CREATE TABLE "recent_grads" (\n"index" INTEGER…
      1  CREATE INDEX "ix_recent_grads_index"ON "recent…
```

4 Basic Query Structure

```
[12]: def read_q(q):
          return pd.read_sql_query(q, conn)
```

```
[13]: q = 'select * from sqlite_master'
      read_q(q)
```

```
[13]:      type                  name       tbl_name  rootpage  \
       0  table         recent_grads  recent_grads         2
       1  index  ix_recent_grads_index  recent_grads         3

                                                       sql
       0  CREATE TABLE "recent_grads" (\n"index" INTEGER…
       1  CREATE INDEX "ix_recent_grads_index"ON "recent…
```

```
[14]: read_q('SELECT * FROM recent_grads')
```

```
[14]:      index  Rank  Major_code                                     Major  \
       0        0     1        2419                     PETROLEUM ENGINEERING
       1        1     2        2416              MINING AND MINERAL ENGINEERING
       2        2     3        2415                    METALLURGICAL ENGINEERING
       3        3     4        2417  NAVAL ARCHITECTURE AND MARINE ENGINEERING
       4        4     5        2405                        CHEMICAL ENGINEERING
       ..     …   …         …                                             …
       168    168   169        3609                                   ZOOLOGY
       169    169   170        5201                      EDUCATIONAL PSYCHOLOGY
       170    170   171        5202                         CLINICAL PSYCHOLOGY
       171    171   172        5203                       COUNSELING PSYCHOLOGY
       172    172   173        3501                             LIBRARY SCIENCE

                     Major_category  Total  Sample_size    Men  Women  ShareWomen  \
       0                Engineering   2339           36   2057    282    0.120564
       1                Engineering    756            7    679     77    0.101852
       2                Engineering    856            3    725    131    0.153037
       3                Engineering   1258           16   1123    135    0.107313
       4                Engineering  32260          289  21239  11021    0.341631
       ..                       …      …            …     …      …           …
       168   Biology & Life Science   8409           47   3050   5359    0.637293
       169  Psychology & Social Work   2854            7    522   2332    0.817099
       170  Psychology & Social Work   2838           13    568   2270    0.799859
       171  Psychology & Social Work   4626           21    931   3695    0.798746
       172                Education   1098            2    134    964    0.877960

             …  Part_time  Full_time_year_round  Unemployed  Unemployment_rate  \
       0      …        270                  1207          37           0.018381
       1      …        170                   388          85           0.117241
       2      …        133                   340          16           0.024096
       3      …        150                   692          40           0.050125
       4      …       5180                 16697        1672           0.061098
       ..     …        …                     …           …                  …
       168    …       2190                  3602         304           0.046320
       169    …        572                  1211         148           0.065112
       170    …        648                  1293         368           0.149048
       171    …        965                  2738         214           0.053621
```

```
172   …          237                    410              87           0.104946
```

```
      Median  P25th   P75th  College_jobs  Non_college_jobs  Low_wage_jobs
0     110000  95000  125000          1534               364            193
1      75000  55000   90000           350               257             50
2      73000  50000  105000           456               176              0
3      70000  43000   80000           529               102              0
4      65000  50000   75000         18314              4440            972
..        …      …       …             …                 …              …
168    26000  20000   39000          2771              2947            743
169    25000  24000   34000          1488               615             82
170    25000  25000   40000           986               870            622
171    23400  19200   26000          2403              1245            308
172    22000  20000   22000           288               338            192

[173 rows x 22 columns]
```

5   Limit

```
[17]: read_q('SELECT * FROM recent_grads LIMIT 10')
```

```
[17]:    index  Rank  Major_code                                       Major  \
      0      0     1        2419                      PETROLEUM ENGINEERING
      1      1     2        2416               MINING AND MINERAL ENGINEERING
      2      2     3        2415                   METALLURGICAL ENGINEERING
      3      3     4        2417  NAVAL ARCHITECTURE AND MARINE ENGINEERING
      4      4     5        2405                       CHEMICAL ENGINEERING
      5      5     6        2418                        NUCLEAR ENGINEERING
      6      6     7        6202                           ACTUARIAL SCIENCE
      7      7     8        5001                   ASTRONOMY AND ASTROPHYSICS
      8      8     9        2414                      MECHANICAL ENGINEERING
      9      9    10        2408                      ELECTRICAL ENGINEERING

           Major_category  Total  Sample_size    Men  Women  ShareWomen  …  \
      0         Engineering   2339           36   2057    282    0.120564  …
      1         Engineering    756            7    679     77    0.101852  …
      2         Engineering    856            3    725    131    0.153037  …
      3         Engineering   1258           16   1123    135    0.107313  …
      4         Engineering  32260          289  21239  11021    0.341631  …
      5         Engineering   2573           17   2200    373    0.144967  …
      6            Business   3777           51    832    960    0.535714  …
      7  Physical Sciences   1792           10   2110   1667    0.441356  …
      8         Engineering  91227         1029  12953   2105    0.139793  …
      9         Engineering  81527          631   8407   6548    0.437847  …

         Part_time  Full_time_year_round  Unemployed  Unemployment_rate  Median  \
```

3

```
0     270              1207        37         0.018381  110000
1     170               388        85         0.117241   75000
2     133               340        16         0.024096   73000
3     150               692        40         0.050125   70000
4    5180             16697      1672         0.061098   65000
5     264              1449       400         0.177226   65000
6     296              2482       308         0.095652   62000
7     553               827        33         0.021167   62000
8   13101             54639      4650         0.057342   60000
9   12695             41413      3895         0.059174   60000

    P25th   P75th  College_jobs  Non_college_jobs  Low_wage_jobs
0   95000  125000          1534               364            193
1   55000   90000           350               257             50
2   50000  105000           456               176              0
3   43000   80000           529               102              0
4   50000   75000         18314              4440            972
5   50000  102000          1142               657            244
6   53000   72000          1768               314            259
7   31500  109000           972               500            220
8   48000   70000         52844             16384           3253
9   45000   72000         45829             10874           3170

[10 rows x 22 columns]
```

6 Specify Column Name

```
[18]: read_q('SELECT ShareWomen FROM recent_grads LIMIT 10')
```

```
[18]:    ShareWomen
      0    0.120564
      1    0.101852
      2    0.153037
      3    0.107313
      4    0.341631
      5    0.144967
      6    0.535714
      7    0.441356
      8    0.139793
      9    0.437847
```

```
[24]: read_q('SELECT ShareWomen,Major FROM recent_grads LIMIT 5')
```

```
[24]:    ShareWomen                           Major
      0    0.120564            PETROLEUM ENGINEERING
      1    0.101852      MINING AND MINERAL ENGINEERING
```

```
2    0.153037            METALLURGICAL ENGINEERING
3    0.107313  NAVAL ARCHITECTURE AND MARINE ENGINEERING
4    0.341631              CHEMICAL ENGINEERING
```

7   Where

- Where in urdu -> jhn pr yay hu
- rows ko filter krny k liyay huta hy

```
[26]: read_q('SELECT ShareWomen, Major FROM recent_grads WHERE ShareWomen > 0.5')
```

```
[26]:     ShareWomen                           Major
      0    0.535714                 ACTUARIAL SCIENCE
      1    0.578766                 COMPUTER SCIENCE
      2    0.558548          ENVIRONMENTAL ENGINEERING
      3    0.896019                          NURSING
      4    0.750473  INDUSTRIAL PRODUCTION TECHNOLOGIES
      ..       …                              …
      92   0.637293                          ZOOLOGY
      93   0.817099            EDUCATIONAL PSYCHOLOGY
      94   0.799859               CLINICAL PSYCHOLOGY
      95   0.798746             COUNSELING PSYCHOLOGY
      96   0.877960                  LIBRARY SCIENCE

      [97 rows x 2 columns]
```

Use ''' ''' or """ """ for multiline query

```
[28]: read_q(''' SELECT ShareWomen, Major
          FROM recent_grads
          WHERE ShareWomen > 0.5
          LIMIT 10''')
```

```
[28]:     ShareWomen                           Major
      0    0.535714                 ACTUARIAL SCIENCE
      1    0.578766                 COMPUTER SCIENCE
      2    0.558548          ENVIRONMENTAL ENGINEERING
      3    0.896019                          NURSING
      4    0.750473  INDUSTRIAL PRODUCTION TECHNOLOGIES
      5    0.707719    COMPUTER AND INFORMATION SYSTEMS
      6    0.526476              INFORMATION SCIENCES
      7    0.753927              APPLIED MATHEMATICS
      8    0.524153                     PHARMACOLOGY
      9    0.688999                     OCEANOGRAPHY
```

8 Operators

`<, >, <=, >=, <>, =, AND, OR, NOT`

```
[29]: read_q(''' SELECT ShareWomen, Major
          FROM recent_grads
          WHERE ShareWomen > 0.5 AND ShareWomen < 0.8''')
```

```
[29]:      ShareWomen                              Major
       0     0.535714                    ACTUARIAL SCIENCE
       1     0.578766                     COMPUTER SCIENCE
       2     0.558548            ENVIRONMENTAL ENGINEERING
       3     0.750473   INDUSTRIAL PRODUCTION TECHNOLOGIES
       4     0.707719      COMPUTER AND INFORMATION SYSTEMS
       ..         …                                    …
       74    0.629505              DRAMA AND THEATER ARTS
       75    0.666119             COMPOSITION AND RHETORIC
       76    0.637293                              ZOOLOGY
       77    0.799859                  CLINICAL PSYCHOLOGY
       78    0.798746                COUNSELING PSYCHOLOGY

       [79 rows x 2 columns]
```

9 Task 1

- return the ShareWomen and Major where either ShareWomen is greater then 0.5 or Share-Women is less than and equal to 0.5 and the majorCategory is Engineering.

```
[31]: read_q("""SELECT ShareWomen, Major
      FROM recent_grads
      WHERE (ShareWomen > 0.5) OR (ShareWomen <= 0.5 AND Major_category ==␣
       ↪'Engineering') """)
```

```
[31]:      ShareWomen                                    Major
       0     0.120564                     PETROLEUM ENGINEERING
       1     0.101852               MINING AND MINERAL ENGINEERING
       2     0.153037                  METALLURGICAL ENGINEERING
       3     0.107313   NAVAL ARCHITECTURE AND MARINE ENGINEERING
       4     0.341631                       CHEMICAL ENGINEERING
       ..         …                                          …
       119   0.637293                                    ZOOLOGY
       120   0.817099                     EDUCATIONAL PSYCHOLOGY
       121   0.799859                        CLINICAL PSYCHOLOGY
       122   0.798746                      COUNSELING PSYCHOLOGY
       123   0.877960                            LIBRARY SCIENCE

       [124 rows x 2 columns]
```

10 Order By

```
[36]: read_q("""SELECT ShareWomen, Major
FROM recent_grads
WHERE (ShareWomen > 0.5) OR (ShareWomen <= 0.5 AND Major_category ==␣
 ↪'Engineering')
ORDER BY ShareWomen""")
```

```
[36]:      ShareWomen                                       Major
       0     0.000000    MISCELLANEOUS ENGINEERING TECHNOLOGIES
       1     0.090713                   ENGINEERING TECHNOLOGIES
       2     0.101852                MINING AND MINERAL ENGINEERING
       3     0.107313  NAVAL ARCHITECTURE AND MARINE ENGINEERING
       4     0.119559                      BIOMEDICAL ENGINEERING
       ..        …                                           …
       119   0.910933                            ANIMAL SCIENCES
       120   0.923745                       ELEMENTARY EDUCATION
       121   0.927807           MATHEMATICS AND COMPUTER SCIENCE
       122   0.967998                   EARLY CHILDHOOD EDUCATION
       123   0.968954                  ANTHROPOLOGY AND ARCHEOLOGY

       [124 rows x 2 columns]
```

```
[37]: read_q("""SELECT ShareWomen, Major
FROM recent_grads
WHERE (ShareWomen > 0.5) OR (ShareWomen <= 0.5 AND Major_category ==␣
 ↪'Engineering')
ORDER BY ShareWomen DESC """)
```

```
[37]:      ShareWomen                                       Major
       0     0.968954                  ANTHROPOLOGY AND ARCHEOLOGY
       1     0.967998                   EARLY CHILDHOOD EDUCATION
       2     0.927807           MATHEMATICS AND COMPUTER SCIENCE
       3     0.923745                       ELEMENTARY EDUCATION
       4     0.910933                            ANIMAL SCIENCES
       ..        …                                           …
       119   0.119559                      BIOMEDICAL ENGINEERING
       120   0.107313  NAVAL ARCHITECTURE AND MARINE ENGINEERING
       121   0.101852                MINING AND MINERAL ENGINEERING
       122   0.090713                   ENGINEERING TECHNOLOGIES
       123   0.000000    MISCELLANEOUS ENGINEERING TECHNOLOGIES

       [124 rows x 2 columns]
```

```
[39]: read_q("""SELECT ShareWomen, Major
FROM recent_grads
```

```
WHERE (ShareWomen > 0.5) OR (ShareWomen <= 0.5 AND Major_category ==␣
 ↪'Engineering')
ORDER BY 1 DESC """)
```

```
[39]:      ShareWomen                                       Major
      0      0.968954              ANTHROPOLOGY AND ARCHEOLOGY
      1      0.967998                 EARLY CHILDHOOD EDUCATION
      2      0.927807          MATHEMATICS AND COMPUTER SCIENCE
      3      0.923745                       ELEMENTARY EDUCATION
      4      0.910933                            ANIMAL SCIENCES
      ..          …                                         …
      119    0.119559                     BIOMEDICAL ENGINEERING
      120    0.107313  NAVAL ARCHITECTURE AND MARINE ENGINEERING
      121    0.101852               MINING AND MINERAL ENGINEERING
      122    0.090713                    ENGINEERING TECHNOLOGIES
      123    0.000000  MISCELLANEOUS ENGINEERING TECHNOLOGIES

      [124 rows x 2 columns]
```

```
[40]: read_q("""SELECT ShareWomen, Major
      FROM recent_grads
      WHERE (ShareWomen > 0.5) OR (ShareWomen <= 0.5 AND Major_category ==␣
       ↪'Engineering')
      ORDER BY 1,2 ASC """)
```

```
[40]:      ShareWomen                                       Major
      0      0.000000     MISCELLANEOUS ENGINEERING TECHNOLOGIES
      1      0.090713                    ENGINEERING TECHNOLOGIES
      2      0.101852               MINING AND MINERAL ENGINEERING
      3      0.107313  NAVAL ARCHITECTURE AND MARINE ENGINEERING
      4      0.119559                     BIOMEDICAL ENGINEERING
      ..          …                                         …
      119    0.910933                            ANIMAL SCIENCES
      120    0.923745                       ELEMENTARY EDUCATION
      121    0.927807          MATHEMATICS AND COMPUTER SCIENCE
      122    0.967998                 EARLY CHILDHOOD EDUCATION
      123    0.968954              ANTHROPOLOGY AND ARCHEOLOGY

      [124 rows x 2 columns]
```

```
[44]: read_q("""SELECT ShareWomen, Major
      FROM recent_grads
      WHERE (ShareWomen > 0.5) OR (ShareWomen <= 0.5 AND Major_category ==␣
       ↪'Engineering')
      ORDER BY 1 DESC,2 ASC
      LIMIT 5""")
```

```
[44]:     ShareWomen                               Major
     0     0.968954          ANTHROPOLOGY AND ARCHEOLOGY
     1     0.967998            EARLY CHILDHOOD EDUCATION
     2     0.927807  MATHEMATICS AND COMPUTER SCIENCE
     3     0.923745                ELEMENTARY EDUCATION
     4     0.910933                     ANIMAL SCIENCES
```

```
[50]: read_q("""SELECT ShareWomen
            FROM recent_grads
            WHERE (ShareWomen > 0.5) OR (ShareWomen <= 0.5 AND Major_category ==␣
      ↪'Engineering')
      """)
```

```
[50]:        ShareWomen
     0          0.120564
     1          0.101852
     2          0.153037
     3          0.107313
     4          0.341631
     ..              …
     119        0.637293
     120        0.817099
     121        0.799859
     122        0.798746
     123        0.877960

     [124 rows x 1 columns]
```

```
[49]: read_q("""SELECT COUNT(ShareWomen)
            FROM recent_grads
            WHERE (ShareWomen > 0.5) OR (ShareWomen <= 0.5 AND Major_category ==␣
      ↪'Engineering')
      """)
```

```
[49]:    COUNT(ShareWomen)
     0                124
```

11 Aliases

```
[53]: read_q("""SELECT COUNT(ShareWomen) as No_of_Eng_Field
            FROM recent_grads
            WHERE Major_category == 'Engineering'
      """)
```

```
[53]:    No_of_Eng_Field
     0                29
```

```
read_q("""SELECT COUNT(ShareWomen) No_of_Eng_Field
        FROM recent_grads
        WHERE Major_category == 'Engineering'
""")
```

[54]:   No_of_Eng_Field
      0   29

```
read_q("""SELECT COUNT(ShareWomen) 'No of Eng Field'
        FROM recent_grads
        WHERE Major_category == 'Engineering'
""")
```

[55]:   No of Eng Field
      0   29

12 Select Distinct

```
read_q("""SELECT DISTINCT Major_category
FROM recent_grads
""")
```

[59]:   Major_category
      0   Engineering
      1   Business
      2   Physical Sciences
      3   Law & Public Policy
      4   Computers & Mathematics
      5   Agriculture & Natural Resources
      6   Industrial Arts & Consumer Services
      7   Arts
      8   Health
      9   Social Science
      10   Biology & Life Science
      11   Education
      12   Humanities & Liberal Arts
      13   Psychology & Social Work
      14   Communications & Journalism
      15   Interdisciplinary

13 GROUP BY

- HAVING is always use in GROUPBY
- But we use WHERE and HAVING in GROUPBY

```
read_q("""SELECT DISTINCT Major_category, avg(ShareWomen)
FROM recent_grads
GROUP BY Major_Category""")
```

```
                      Major_category  avg(ShareWomen)
0          Agriculture & Natural Resources         0.617938
1                                  Arts         0.561851
2                   Biology & Life Science         0.584518
3                               Business         0.405063
4            Communications & Journalism         0.643835
5                Computers & Mathematics         0.512752
6                              Education         0.674986
7                            Engineering         0.257158
8                                 Health         0.616857
9                 Humanities & Liberal Arts         0.676193
10   Industrial Arts & Consumer Services         0.449351
11                        Interdisciplinary         0.495397
12                    Law & Public Policy         0.335990
13                        Physical Sciences         0.508749
14                 Psychology & Social Work         0.777763
15                          Social Science         0.539067
```

```
read_q("""SELECT DISTINCT Major_category, avg(ShareWomen)
FROM recent_grads
GROUP BY 1""")
```

```
                      Major_category  avg(ShareWomen)
0          Agriculture & Natural Resources         0.617938
1                                  Arts         0.561851
2                   Biology & Life Science         0.584518
3                               Business         0.405063
4            Communications & Journalism         0.643835
5                Computers & Mathematics         0.512752
6                              Education         0.674986
7                            Engineering         0.257158
8                                 Health         0.616857
9                 Humanities & Liberal Arts         0.676193
10   Industrial Arts & Consumer Services         0.449351
11                        Interdisciplinary         0.495397
12                    Law & Public Policy         0.335990
13                        Physical Sciences         0.508749
14                 Psychology & Social Work         0.777763
15                          Social Science         0.539067
```

14 HAVING

- HAVING is only put in these type of functions like: avg, sum, count

```
[74]: read_q("""SELECT DISTINCT Major_category, avg(ShareWomen) AVG_SW
FROM recent_grads
GROUP BY 1
HAVING AVG_SW > 0.5""")
```

```
[74]:                Major_category     AVG_SW
    0   Agriculture & Natural Resources  0.617938
    1                            Arts    0.561851
    2              Biology & Life Science  0.584518
    3       Communications & Journalism  0.643835
    4          Computers & Mathematics   0.512752
    5                       Education    0.674986
    6                          Health    0.616857
    7          Humanities & Liberal Arts  0.676193
    8                Physical Sciences    0.508749
    9          Psychology & Social Work  0.777763
    10               Social Science     0.539067
```

```
[75]: read_q("""SELECT DISTINCT Major_category, avg(ShareWomen) AVG_SW
FROM recent_grads
WHERE Major_category != 'Social Science'
GROUP BY 1
HAVING AVG_SW > 0.5""")
```

```
[75]:                Major_category     AVG_SW
    0   Agriculture & Natural Resources  0.617938
    1                            Arts    0.561851
    2              Biology & Life Science  0.584518
    3       Communications & Journalism  0.643835
    4          Computers & Mathematics   0.512752
    5                       Education    0.674986
    6                          Health    0.616857
    7          Humanities & Liberal Arts  0.676193
    8                Physical Sciences    0.508749
    9          Psychology & Social Work  0.777763
```

```
[78]: read_q("""SELECT DISTINCT Major_category
FROM recent_grads
WHERE MAjor_category != 'Social Science'
""")
```

```
[78]:                Major_category
    0                     Engineering
    1                        Business
    2               Physical Sciences
    3             Law & Public Policy
    4         Computers & Mathematics
```

```
 5           Agriculture & Natural Resources
 6     Industrial Arts & Consumer Services
 7                                     Arts
 8                                   Health
 9                    Biology & Life Science
10                                Education
11              Humanities & Liberal Arts
12               Psychology & Social Work
13             Communications & Journalism
14                        Interdisciplinary
```

[83]:
```
read_q("""SELECT avg(ShareWomen) AVG, Major_category
FROM recent_grads
WHERE Major_category != 'Business'
GROUP BY 2
HAVING AVG < 0.5
""")
```

[83]:
```
        AVG                  Major_category
0  0.257158                       Engineering
1  0.449351  Industrial Arts & Consumer Services
2  0.495397                  Interdisciplinary
3  0.335990                Law & Public Policy
```

[ ]: