

FE512 Final Project Data Import 2

In [58]: `%load_ext sql`

The sql extension is already loaded. To reload it, use:
`%reload_ext sql`

In [59]: `%sql mysql+pymysql://root:@fe512_mysql/fe512db`

Out[59]: 'Connected: root@fe512db'

In [64]: `%%sql CREATE DATABASE hint`

* mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.

Out[64]: []

In [5]: `%sql USE hint`

* mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.

Out[5]: []

```
In [66]: %sql SHOW TABLES
```

```
* mysql+pymysql://root:***@fe512_mysql/fe512db
19 rows affected.
```

Out[66]:

Tables_in_hint
Cancer_Cat_Value_Label
Race_Cat2_Value_Label
cancer_info
prostate_analysis
section_a
section_b
section_c
section_d
section_e
section_f
section_g
section_h
section_i
section_j
section_k
section_l
section_m
section_n
section_o

section_o

```
In [9]: %%sql
DROP TABLE IF EXISTS o_response; #db name
CREATE TABLE IF NOT EXISTS section_o(
    PersonID VARCHAR(20),
    Age INT,
    OccupationStatus INT,
    OccupationStatus_OS VARCHAR(255),
    Employed INT,
    Unemployed INT,
    Homemaker INT,
    Student INT,
    Retired INT,
    Disabled INT,
    OtherOcc INT,
    MultiOcc INT,
    ActiveDutyArmedForces INT,
    ReceivedCareVA INT,
    MaritalStatus INT,
    Education INT,
    BornInUSA INT,
    YearCameToUSA INT,
    SpeakEnglish INT,
    NotHisp INT,
    Mexican INT,
    PuertoRican INT,
    Cuban INT,
    OthHisp INT,
    Hisp_Cat INT,
    White INT,
    Black INT,
    AmerInd INT,
    AsInd INT,
    Chinese INT,
    Filipino INT,
    Japanese INT,
    Korean INT,
    Vietnamese INT,
    OthAsian INT,
    Hawaiian INT,
    Guamanian INT,
    Samoan INT,
    OthPacIsl INT,
    Race_Cat2 INT,
    SexualOrientation INT,
    SexualOrientation_OS VARCHAR(255),
    TotalHousehold INT,
    R_HHAdults INT,
    HHAdults_Num INT,
    SelfGender INT,
    SelfAge INT,
    SelfMOB INT,
    HHAdultGender2 INT,
    HHAdultAge2 INT,
    HHAdultMOB2 INT,
    HHAdultGender3 INT,
    HHAdultAge3 INT,
```

```

HHAdultMOB3 INT,
HHAdultGender4 INT,
HHAdultAge4 INT,
HHAdultMOB4 INT,
HHAdultGender5 INT,
HHAdultAge5 INT,
HHAdultMOB5 INT,
ChildrenInHH INT,
RentOrOwn INT,
CellPhone INT,
PhoneInHome INT,
IncomeRanges INT,
MailSurveyTime_Min INT,
MailSurveyTime_Hrs INT,
TypeOfAddressA INT,
TypeOfAddressB INT,
TypeOfAddressC INT,
TypeOfAddressD INT
);

```

```

* mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
0 rows affected.

```

Out[9]: []

```

In [67]: %%sql
SELECT count(*)
FROM information_schema.columns
WHERE table_name = 'section_o';

```

```

* mysql+pymysql://root:***@fe512_mysql/fe512db
1 rows affected.

```

Out[67]:

count(*)
71

```

In [14]: %%sql SHOW TABLES

```

```

* mysql+pymysql://root:***@fe512_mysql/fe512db
2 rows affected.

```

Out[14]:

Tables_in_hint
section_n
section_o

In [11]: %%sql

```
LOAD DATA
  INFILE '/home/data/section_o.csv'
  INTO TABLE section_o
  FIELDS
    TERMINATED BY ','
    OPTIONALLY ENCLOSED BY '"'
    ESCAPED BY '\\'
  LINES
    TERMINATED BY '\n'
  IGNORE 1 LINES;
```

* mysql+pymysql://root:***@fe512_mysql/fe512db
3285 rows affected.

Out[11]: []

In [12]: %%sql SELECT * FROM section_o LIMIT 1

* mysql+pymysql://root:***@fe512_mysql/fe512db
1 rows affected.

Out[12]:

PersonID	Age	OccupationStatus	OccupationStatus_OS	Employed	Unemployed	Home
60000001-02	49	3	-1	2	2	1

section_n

In [20]: %%sql

```
DROP TABLE IF EXISTS section_n; #db name
CREATE TABLE IF NOT EXISTS section_n(
  PersonID VARCHAR(20),
  ChanceGetCancer INT,
  EverythingCauseCancer INT,
  PreventNotPossible INT,
  TooManyRecommendations INT,
  CancerMoreCommon INT,
  CancerFatal INT,
  RatherNotKnowChance INT,
  FreqWorryCancer INT,
  FamilyEverHadCancer INT
);
```

* mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
0 rows affected.

Out[20]: []

In [21]: %%sql

```
LOAD DATA
  INFILE '/home/data/section_n.csv'
  INTO TABLE section_n
  FIELDS
    TERMINATED BY ','
    OPTIONALLY ENCLOSED BY '"'
    ESCAPED BY '\\'
  LINES
    TERMINATED BY '\n'
  IGNORE 1 LINES;
```

* mysql+pymysql://root:***@fe512_mysql/fe512db
3285 rows affected.

Out[21]: []

In [22]: %sql SELECT * FROM section_n LIMIT 1;

* mysql+pymysql://root:***@fe512_mysql/fe512db
1 rows affected.

Out[22]:

PersonID	ChanceGetCancer	EverythingCauseCancer	PreventNotPossible	TooManyRe
60000001-02	5	2	3	3

section_m

In [23]: %%sql

```

DROP TABLE IF EXISTS section_m; #db name
CREATE TABLE IF NOT EXISTS section_m(
    PersonID VARCHAR(20),
    EverHadCancer INT,
    CaBladder INT,
    CaBone INT,
    CaBreast INT,
    CaCervical INT,
    CaColon INT,
    CaEndometrial INT,
    CaHeadNeck INT,
    CaHodgkins INT,
    CaLeukemia INT,
    CaLiver INT,
    CaLung INT,
    CaMelanoma INT,
    CaNonHodgkin INT,
    CaOral INT,
    CaOvarian INT,
    CaPancreatic INT,
    CaPharyngeal INT,
    CaProstate INT,
    CaRectal INT,
    CaRenal INT,
    CaSkin INT,
    CaStomach INT,
    CaOther INT,
    CaOther_OS VARCHAR(255),
    Cancer_Cat INT,
    WhenDiagnosedCancer INT,
    UndergoCancerTreatment INT,
    CancerTx_Chemo INT,
    CancerTx_Radiation INT,
    CancerTx_Surgery INT,
    CancerTx_Other INT,
    HowLongFinishTreatment_Cat INT,
    CancerTxSummary INT,
    CancerDeniedCoverage INT,
    CancerHurtFinances INT,
    CancerAbilityToWork INT,
    ClinicalTrialCancerTx INT,
    DiscussedClinicalTrial INT
);

```

```

* mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
0 rows affected.

```

```

/opt/conda/lib/python3.6/site-packages/pymysql/cursors.py:170: Warning:
(1051, "Unknown table 'hint.section_m'")
  result = self._query(query)

```

Out[23]: []

In [24]: **%sql** SHOW TABLES;

* mysql+pymysql://root:***@fe512_mysql/fe512db
3 rows affected.

Out[24]:

Tables_in_hint
section_m
section_n
section_o

In [25]: **%%sql**

```
LOAD DATA
  INFILE '/home/data/section_m.csv'
  INTO TABLE section_m
  FIELDS
    TERMINATED BY ','
    OPTIONALLY ENCLOSED BY '"'
    ESCAPED BY '\\'
  LINES
    TERMINATED BY '\n'
  IGNORE 1 LINES;
```

* mysql+pymysql://root:***@fe512_mysql/fe512db
3285 rows affected.

Out[25]: []

In [26]: **%sql** SELECT * FROM section_m LIMIT 1;

* mysql+pymysql://root:***@fe512_mysql/fe512db
1 rows affected.

Out[26]:

PersonID	EverHadCancer	CaBladder	CaBone	CaBreast	CaCervical	CaColon	CaEndo
60000001-02	2	-1	-1	-1	-1	-1	-1

section_l


```
In [27]: %%sql
DROP TABLE IF EXISTS section_1; #db name
CREATE TABLE IF NOT EXISTS section_1(
    PersonID VARCHAR(20),
    HeardHPV INT,
    HPVCauseCancer_Cervical INT,
    HPVCauseCancer_Penile INT,
    HPVCauseCancer_Anal INT,
    HPVCauseCancer_Oral INT,
    HPVSTD INT,
    HPVMedicalTreatment INT,
    HeardHPVVaccine2 INT,
    HPVShotPrevent INT,
    FamBetween9and27 INT,
    DiscussHPVVaccination12m INT,
    RecommendHPVShot INT
);
```

```
* mysql+pymysql://root:***@fe512_mysql/fe512db
```

```
0 rows affected.
```

```
0 rows affected.
```

```
/opt/conda/lib/python3.6/site-packages/pymysql/cursors.py:170: Warning:
(1051, "Unknown table 'hint.section_1'")
```

```
result = self._query(query)
```

```
Out[27]: []
```

```
In [28]: %%sql SHOW TABLES;
```

```
* mysql+pymysql://root:***@fe512_mysql/fe512db
```

```
4 rows affected.
```

```
Out[28]:
```

Tables_in_hint
section_l
section_m
section_n
section_o

In [29]: %%sql

```
LOAD DATA
  INFILE '/home/data/section_1.csv'
  INTO TABLE section_1
  FIELDS
    TERMINATED BY ','
    OPTIONALLY ENCLOSED BY '"'
    ESCAPED BY '\\'
  LINES
    TERMINATED BY '\n'
  IGNORE 1 LINES;
```

* mysql+pymysql://root:***@fe512_mysql/fe512db
3285 rows affected.

Out[29]: []

In [30]: %sql SELECT * FROM section_1 LIMIT 1;

* mysql+pymysql://root:***@fe512_mysql/fe512db
1 rows affected.

Out[30]:

PersonID	HeardHPV	HPVCauseCancer_Cervical	HPVCauseCancer_Penile	HPVCauseC
60000001-02	1	1	1	1

section_k

In [37]: %%sql

```
DROP TABLE IF EXISTS section_k; #db name
CREATE TABLE IF NOT EXISTS section_k(
  PersonID VARCHAR(20),
  GenderC INT,
  WhenPapTest INT,
  WhenMammogram INT,
  EverHadPSATest INT
);
```

* mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
0 rows affected.

Out[37]: []

In [38]: %sql SHOW TABLES;

* mysql+pymysql://root:***@fe512_mysql/fe512db
5 rows affected.

Out[38]:

Tables_in_hint
section_k
section_l
section_m
section_n
section_o

In [39]: %%sql

```
LOAD DATA
  INFILE '/home/data/section_k.csv'
  INTO TABLE section_k
  FIELDS
    TERMINATED BY ','
    OPTIONALLY ENCLOSED BY '"'
    ESCAPED BY '\\'
  LINES
    TERMINATED BY '\n'
  IGNORE 1 LINES;
```

* mysql+pymysql://root:***@fe512_mysql/fe512db
3285 rows affected.

Out[39]: []

In [40]: %sql SELECT * FROM section_k LIMIT 1;

* mysql+pymysql://root:***@fe512_mysql/fe512db
1 rows affected.

Out[40]:

PersonID	GenderC	WhenPapTest	WhenMammogram	EverHadPSATest
60000001-02	2	4	3	-1

section_j

```
In [41]: %%sql
DROP TABLE IF EXISTS section_j; #db name
CREATE TABLE IF NOT EXISTS section_j(
    PersonID VARCHAR(20),
    Smoke100 INT,
    SmokeNow INT,
    TriedQuit INT,
    ConsiderQuit INT,
    ElectCigLessHarm INT,
    UsedECigEver INT,
    UseECigNow INT,
    DrTalkLungTest INT,
    SmokelessLessHarm INT,
    HookahLessHarm INT
);

* mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
0 rows affected.

/opt/conda/lib/python3.6/site-packages/pymysql/cursors.py:170: Warning:
(1051, "Unknown table 'hint.section_j'")
result = self._query(query)
```

Out[41]: []

```
In [42]: %sql SHOW TABLES;

* mysql+pymysql://root:***@fe512_mysql/fe512db
6 rows affected.
```

Out[42]:

Tables_in_hint
section_j
section_k
section_l
section_m
section_n
section_o

In [45]: %%sql

```
LOAD DATA
  INFILE '/home/data/section_j.csv'
  INTO TABLE section_j
  FIELDS
    TERMINATED BY ','
    OPTIONALLY ENCLOSED BY '"'
    ESCAPED BY '\\'
  LINES
    TERMINATED BY '\n'
  IGNORE 1 LINES;
```

```
* mysql+pymysql://root:***@fe512_mysql/fe512db
3285 rows affected.
```

Out[45]: []

In [47]: %sql SELECT * FROM section_j LIMIT 1;

```
* mysql+pymysql://root:***@fe512_mysql/fe512db
1 rows affected.
```

Out[47]:

PersonID	Smoke100	SmokeNow	TriedQuit	ConsiderQuit	ElectCigLessHarm	UsedECi
60000001-02	1	3	-1	-1	2	2

section_i

In [48]: %%sql

```
DROP TABLE IF EXISTS section_i; #db name
CREATE TABLE IF NOT EXISTS section_i(
  PersonID VARCHAR(20),
  TimesModerateExercise INT,
  HowLongModerateExerciseMn INT,
  HowLongModerateExerciseHr INT,
  TimesStrengthTraining INT,
  TanningBed INT,
  SkinCancerHPEexam INT,
  SkinCancerSelfCheck INT
);
```

```
* mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
0 rows affected.
```

```
/opt/conda/lib/python3.6/site-packages/pymysql/cursors.py:170: Warning:
(1051, "Unknown table 'hint.section_i'")
  result = self._query(query)
```

Out[48]: []

In [49]: **%sql** SHOW TABLES;

```
* mysql+pymysql://root:***@fe512_mysql/fe512db
7 rows affected.
```

Out[49]:

Tables_in_hint
section_i
section_j
section_k
section_l
section_m
section_n
section_o

In [50]: **%%sql**

```
LOAD DATA
  INFILE '/home/data/section_i.csv'
  INTO TABLE section_i
  FIELDS
    TERMINATED BY ','
    OPTIONALLY ENCLOSED BY '"'
    ESCAPED BY '\\'
  LINES
    TERMINATED BY '\n'
  IGNORE 1 LINES;
```

```
* mysql+pymysql://root:***@fe512_mysql/fe512db
3285 rows affected.
```

Out[50]: []

In [51]: **%sql** SELECT * FROM section_i LIMIT 1;

```
* mysql+pymysql://root:***@fe512_mysql/fe512db
1 rows affected.
```

Out[51]:

PersonID	TimesModerateExercise	HowLongModerateExerciseMn	HowLongModerateE
60000001-02	4	45	0

section_h

```
In [53]: %%sql
DROP TABLE IF EXISTS section_h; #db name
CREATE TABLE IF NOT EXISTS section_h(
    PersonID VARCHAR(20),
    UseMenuCalorieInfo INT,
    Fruit INT,
    Vegetables INT,
    AlcoholConditions_Cancer INT,
    AlcoholConditions_HeartDisease INT,
    AlcoholConditions_Diabetes INT,
    AlcoholConditions_Cholesterol INT,
    AlcoholConditions_LiverDisease INT,
    AlcoholConditions_Overweight INT,
    AlcoholIncreaseCancer INT,
    AlcoholReduceHeart INT
);
```

```
* mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
0 rows affected.
```

```
/opt/conda/lib/python3.6/site-packages/pymysql/cursors.py:170: Warning:
(1051, "Unknown table 'hint.section_h'")
  result = self._query(query)
```

Out[53]: []

```
In [54]: %sql SHOW TABLES;
```

```
* mysql+pymysql://root:***@fe512_mysql/fe512db
8 rows affected.
```

Out[54]:

Tables_in_hint
section_h
section_i
section_j
section_k
section_l
section_m
section_n
section_o

In [55]: %%sql

```

LOAD DATA
  INFILE '/home/data/section_h.csv'
  INTO TABLE section_h
  FIELDS
    TERMINATED BY ','
    OPTIONALLY ENCLOSED BY '"'
    ESCAPED BY '\\'
  LINES
    TERMINATED BY '\n'
  IGNORE 1 LINES;

```

```

* mysql+pymysql://root:***@fe512_mysql/fe512db
3285 rows affected.

```

Out[55]: []

In [56]: %sql SELECT * FROM section_h LIMIT 1;

```

* mysql+pymysql://root:***@fe512_mysql/fe512db
1 rows affected.

```

Out[56]:

PersonID	UseMenuCalorieInfo	Fruit	Vegetables	AlcoholConditions_Cancer	AlcoholCo
60000001-02	2	2	3	1	-9

survey_section

In [98]: %%sql

```

DROP TABLE IF EXISTS survey_section;
CREATE TABLE IF NOT EXISTS survey_section(
  sectionID CHAR(1),
  section_description VARCHAR(255)
);

```

```

* mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
0 rows affected.

```

Out[98]: []

In [99]: %sql SHOW TABLES

```
* mysql+pymysql://root:***@fe512_mysql/fe512db
20 rows affected.
```

Out[99]:

Tables_in_hint
Cancer_Cat_Value_Label
Race_Cat2_Value_Label
cancer_info
prostate_analysis
section_a
section_b
section_c
section_d
section_e
section_f
section_g
section_h
section_i
section_j
section_k
section_l
section_m
section_n
section_o
survey_section

In [100]: %%sql

```
LOAD DATA
  INFILE '/home/data/survey_section.csv'
  INTO TABLE survey_section
  FIELDS
    TERMINATED BY ','
    OPTIONALLY ENCLOSED BY '"'
    ESCAPED BY '\\'
  LINES
    TERMINATED BY '\n'
  IGNORE 1 LINES;
```

```
* mysql+pymysql://root:***@fe512_mysql/fe512db
15 rows affected.
```

Out[100]: []

In [101]: %sql SELECT * FROM survey_section

```
* mysql+pymysql://root:***@fe512_mysql/fe512db
15 rows affected.
```

Out[101]:

sectionID	section_description
A	Looking For Health Information
B	Using the Internet to Find Information
C	Your Health Care
D	Medical Records
E	Caregiving
F	Medical Research
G	Your Overall Health
H	Health and Nutrition
I	Physical Activity Exercise and UV Exposure
J	Tobacco Products
K	Screening For Cancer
L	HPV Awareness
M	Your Cancer History
N	Beliefs About Cancer
O	You and Your Household

```
In [102]: %%sql
ALTER TABLE survey_section
ADD PRIMARY KEY (sectionID);

* mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
```

Out[102]: []

```
In [95]: %sql DESCRIBE survey_section;

* mysql+pymysql://root:***@fe512_mysql/fe512db
2 rows affected.
```

Out[95]:

Field	Type	Null	Key	Default	Extra
sectionID	char(1)	NO	PRI	None	
section_description	varchar(255)	YES		None	

section_question_answer

```
In [113]: %%sql
DROP TABLE IF EXISTS section_question_answer;
CREATE TABLE IF NOT EXISTS section_question_answer(
    sectionID CHAR(1),
    answer_abbr VARCHAR(45),
    questionID VARCHAR(5),
    question_description VARCHAR(2000)
);
ALTER TABLE section_question_answer
ADD PRIMARY KEY (answer_abbr);

* mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
0 rows affected.
0 rows affected.
```

Out[113]: []

```
In [119]: %sql DESCRIBE section_question_answer;

* mysql+pymysql://root:***@fe512_mysql/fe512db
4 rows affected.
```

Out[119]:

Field	Type	Null	Key	Default	Extra
sectionID	char(1)	YES		None	
answer_abbr	varchar(45)	NO	PRI	None	
questionID	varchar(5)	YES		None	
question_description	varchar(2000)	YES		None	

```
In [122]: %%sql
LOAD DATA
  INFILE '/home/data/section_question_answer.csv'
  INTO TABLE section_question_answer
  FIELDS
    TERMINATED BY ','
  OPTIONALLY ENCLOSED BY '"'
  ESCAPED BY '\\'
  LINES
    TERMINATED BY '\r\n'
  IGNORE 1 LINES
  (@sectionID, @answer_abbr, @questionID, @question_description) #col
  CSV
  SET sectionID = @sectionID,
      answer_abbr = @answer_abbr,
      questionID = NULLIF(@questionID, ''),
      question_description = @question_description;

* mysql+pymysql://root:***@fe512_mysql/fe512db
337 rows affected.
```

Out[122]: []

```
In [124]: %sql SELECT * FROM section_question_answer LIMIT 10;

* mysql+pymysql://root:***@fe512_mysql/fe512db
10 rows affected.
```

Out[124]:

sectionID	answer_abbr	questionID	question_description
O	385 SelfAge	O14	Self: Age.
O	386 SelfMOB	O14	Self: Month Born
O	390 HHAdultGender2	O14	Adult 2: Gender
D	AccessedFamRec_MyPwd	D16a	Used a login and password assigned to me to access their record.
D	AccessedFamRec_TheirPwd	D16a	Used family members login and password.
D	AccessFamilyMedRec	D15	How many times access a family members online medical record through a secure website/app?
D	AccessOnlineRecord	D4	How many times did you access your online medical record in the last 12 months?
O	ActiveDutyArmedForces	O3	Have you ever served on active duty in the U.S. Armed Forces/Military Reserves/National Guard?
O	Age	O1	What is your age?
H	AlcoholConditions_Cancer	H4a	Drinking too much alcohol - Cancer

person

```
In [125]: %%sql
DROP TABLE IF EXISTS person;
CREATE TABLE IF NOT EXISTS person(
    personID VARCHAR(20)
);
ALTER TABLE person
ADD PRIMARY KEY (personID);

* mysql+pymysql://root:***@fe512_mysql/fe512db
0 rows affected.
0 rows affected.
0 rows affected.

/opt/conda/lib/python3.6/site-packages/pymysql/cursors.py:170: Warning:
(1051, "Unknown table 'hint.person'")
result = self._query(query)
```

Out[125]: []

```
In [126]: %%sql
LOAD DATA
    INFILE '/home/data/person.csv'
    INTO TABLE person
    FIELDS
        TERMINATED BY ','
    OPTIONALLY ENCLOSED BY '"'
        ESCAPED BY '\\'
    LINES
        TERMINATED BY '\r\n'
    IGNORE 1 LINES;

* mysql+pymysql://root:***@fe512_mysql/fe512db
3285 rows affected.
```

Out[126]: []