Elixir- DMDD Assignment 3

Project Description:

This project mainly provides details of all the Diseases, Drugs, Hospitals and the Diagnosis details of the patient. It includes data gathering from various data sources. The gathered data is preprocessed according to the database structure and schema. A detail audit of the data sources is made.

Audit:

1. Diagnosis Data:

Relevancy: The data should meet the requirements for the intended use.

Comments: Data created by HealthMan – a not for profit organisation dedicated to manage the functioning of Hospitals in a professional and optimal manner.

Completeness: The data should not have missing values or miss data records.

Comments:

Bed Grade 113 City_Code_Patient 4532

Above are null value counts.

Diagosis - Filling with "null" value

Bed Grade - Filling with mean value of bed grade

Consistency: the data should have the data format as expected and can be cross reference-able with the same results.

Comments:

Age and Stay have interval values

Diagnosis - Mean of the interval taken

2. Hospitals

Relevancy: The data should meet the requirements for the intended use.

Comments: This dataset is provided by the Homeland Infrastructure Foundation-Level Data (HIFLD) without a licence and for Public Use.

Completeness: the data should not have missing values or miss data records.

Comments:

TTL STAFF

Have constant value -999.

Diagnosis - We have dropped the column as there is no information in TTL STAFF

WEBSITE 377 ARE NOT AVAILABLE.

Diagnosis - Mode Imputation will lead to wrong websites to hospitals. So filled null values with empty string.

Consistency: The data should have the data format as expected and can be cross reference-able with the same results.

Comments:

Data format is as need to directly join with diagnosis data. No preprocessing is done.

3. Drugs:

Relevancy: The data should meet the requirements for the intended use.

Comments:

This dataset is created from web scraping of various health websites

Completeness: the data should not have missing values or miss data records.

Comments:

There are no missing values in diseases-drugs data.

Consistency: The data should have the data format as expected and can be cross reference-able with the same results.

Comments: Data format is as need to directly join with diagnosis data. No preprocessing is done.

Use Cases:

1. Which city needs more health care attention?

SELECT max(cases_count) as cases_count, cityCodePatient from (SELECT count(caseId) as cases_count, cityCodePatient from Diagnosis group by cityCodePatient);

2. List of patients who were referred to multiple hospitals for a given city.

SELECT count(DISTINCT Diagnosis.hospitalCode) as hospital_count, Diagnosis.patientid, Patients.name, Hospitals.city FROM Diagnosis JOIN Patients on Patients.patientid=Diagnosis.patientid JOIN Hospitals on Hospitals.hospitalCode=Diagnosis.hospitalCode
GROUP by Diagnosis.patientid HAVING hospital count>3 AND Hospitals.city='tillamook';

3. Prediction of disease based on a particular symptom

SELECT DISTINCT disease FROM Diseases WHERE symptoms like '%abdominal pain%';

4. What is the most prevalent disease in the city and what are its symptoms?

SELECT max(max_cases) as cases_number, disease_name, symptoms
FROM (SELECT count(caseId) as max_cases, Diagnosis.disease_id as disease_id, Diseases.disease
as disease_name, symptoms
FROM Diagnosis JOIN Diseases on Diseases.disease_id=Diagnosis.disease_id
WHERE hospitalCode in (SELECT hospitalCode FROM Hospitals WHERE city like '%baltimore%')
GROUP by Diagnosis.disease_id);

5. Identify the list of all the patients less than a particular age affected by a particular disease in extreme severity and urgent admission.

SELECT DISTINCT Patients.patientid, Patients.name FROM Diagnosis as dgs JOIN Diseases as dis on dis.disease_id=dgs.disease_id

JOIN Patients on Patients.patientid=dgs.patientid WHERE dis.disease like '%bipolar disorder%' AND severityOfIllness='Extreme' AND typeOfAdmission='Urgent' AND age<30;

6. Total revenue generated by a particular hospital with admission deposits on a particular date.

SELECT dgs.hospitalCode, hos.name, sum(admissionDeposit) as totalAdmissionDeposit FROM Diagnosis as dgs JOIN Hospitals as hos on hos.hospitalCode=dgs.hospitalCode WHERE startDate='2022-09-28' AND dgs.hospitalCode=28;

7. List of patients who underwent diagnosis in a particular hospital for a given department

SELECT DISTINCT Patients.patientid, Patients.name FROM Diagnosis JOIN Patients on Patients.patientid=Diagnosis.patientid
WHERE startDate BETWEEN '2022-04-01' AND '2022-05-01' AND hospitalCode=19 AND department='radiotherapy';

8. Number of case registrations for all the hospitals in a given duration.

SELECT count(caseId) as registrations_count, Diagnosis.hospitalCode, name FROM Diagnosis JOIN Hospitals on Hospitals.hospitalCode=Diagnosis.hospitalCode WHERE startDate BETWEEN '2021-03-01' AND '2021-04-01' GROUP by Diagnosis.hospitalCode;

9. List of patients whose length of stay is maximum in a given hospital along with age and name of the patient.

SELECT DISTINCT disease, Patients.name, age FROM Diagnosis JOIN Diseases on Diseases.disease_id=Diagnosis.disease_id

JOIN Patients on Patients.patientid=Diagnosis.patientid WHERE stay=120 AND hospitalCode=5;

10. Hospital details where the available beds are greater than 25

SELECT name, beds, telephone, website, zipcode FROM Hospitals WHERE beds>25;

Datasources:

Medicine_prescription_records:

https://www.kaggle.com/datasets/manncodes/drug-prescription-to-disease-dataset

Hospitals:

https://www.kaggle.com/datasets/carlosaguayo/usa-hospitals

diseases to drug:

https://www.kaggle.com/datasets/manncodes/drug-prescription-to-disease-dataset?resource=download

patients disease (diagnosis):

https://datahack.analyticsvidhya.com/contest/janatahack-healthcare-analytics-ii/

Diagnosis:

https://datahack.analyticsvidhya.com/contest/janatahack-healthcare-analytics-ii/#ProblemStatement

Create and Insert Queries:

Create Statements:

```
CREATE TABLE "Diagnosis" (
      "caseld"
                    INTEGER,
      "hospitalCode"INTEGER,
      "department" TEXT,
      "wardType"
                    TEXT,
      "wardFacilityCode"
                          TEXT,
      "bedGrade"
                    REAL.
                    INTEGER,
      "patientid"
      "cityCodePatient"
                          REAL,
      "typeOfAdmission"
                          TEXT,
      "severityOfIllness"
                          TEXT,
      "visitorsWithPatient" INTEGER,
      "admissionDeposit"
                          REAL,
      "stay" REAL,
      "startDate"
                    TEXT,
      "endDate"
                    TEXT.
      "disease_id" INTEGER
);
CREATE TABLE "Diseases" (
      "disease"
                    TEXT,
      "symptoms"
                    TEXT,
      "disease id"
                    INTEGER
);
CREATE TABLE "Drugs" (
      "drug" TEXT,
      "drug id"
                    INTEGER.
      "disease_id" INTEGER
);
CREATE TABLE "Hospitals" (
      "hospitalCode"INTEGER,
      "name"TEXT,
      "city" TEXT,
      "zipcode"
                    TEXT,
      "beds" TEXT,
      "telephone"
                    TEXT,
      "type" TEXT,
      "status"
                    TEXT,
      "population"
                    INTEGER,
      "county"
                    TEXT,
      "website"
                    TEXT,
      "helipad"
                    TEXT
);
CREATE TABLE "Patients" (
      "patientid"
                    INTEGER,
      "name"TEXT,
      "age" REAL
```

```
);
CREATE TABLE "Tags" (
      "tweet_id"
                   TEXT,
      "tags" TEXT
);
CREATE TABLE "Tweets" (
      "tweet_id"
                   TEXT,
      "twitter_handle"
                          TEXT,
      "tweet text" TEXT,
      "created_at" TIMESTAMP,
      "retweet_count"
                          INTEGER,
      "likes_count" INTEGER
);
CREATE TABLE "Twitter_User" (
      "user_id"
                   TEXT,
      "user_handle" TEXT,
      "user_name" TEXT,
      "user_profile_image" TEXT,
      "description" TEXT,
      "followers_count"
                          INTEGER,
      "created_at" TIMESTAMP
);
CREATE TABLE "hospitals_sentiment_by_disease" (
      "Tweet Text" TEXT,
      "sentiment"
                    REAL,
      "hospital"
                   TEXT,
      "disease"
                    TEXT
);
CREATE TABLE "medicines_sentiment_by_disease" (
      "Tweet Text" TEXT,
      "sentiment"
                    REAL,
      "medicine"
                    TEXT,
      "disease"
                    TEXT
);
CREATE TABLE "query_sentiment" (
      "tweet_id"
                   TEXT,
      "query"TEXT,
      "sentiment"
                   REAL,
      "type" TEXT
);
```

Insert Statements

Diagnosis table:

INSERT INTO "main". "Diagnosis"

("caseId","hospitalCode","department","wardType","wardFacilityCode","bedGrade","patientid","cityCodePatient ","typeOfAdmission","severityOfIllness","visitorsWithPatient","admissionDeposit","stay","startDate","endDate"," disease_id") VALUES

("1","8","radiotherapy","R","F","2.0","31397","7","Emergency","Extreme","2","4911","5","2022-03-05","2022-03-08","31");

Diseases table:

INSERT INTO "main". "Diseases" ("disease", "symptoms", "disease_id") VALUES ('vulvodynia', 'pelvic pain, sharp abdominal pain, lower abdominal pain', '201');

Drugs table:

INSERT INTO "main". "Drugs" ("drug", "drug_id", "disease_id") VALUES ("clonazepam", "145", "134");

Hospitals table:

INSERT INTO

"main"."Hospitals"("hospitalCode","name","city","zipcode","beds","telephone","type","status","population","county","website","helipad") VALUES ("0","east jefferson general hospital","metairie","70006","420.0","(504) 454-4000","GENERAL ACUTE CARE","OPEN","420","JEFFERSON","http://www.eigh.org","Y");

Patients table:

INSERT INTO "main". "Patients" ("patientid", "name", "age") VALUES ("1", "Jack", "42");

Tags table:

INSERT INTO "main"."Tags"("tweet_id","tags") VALUES ("1597781304250028032","weightloss");

Tweets table:

INSERT INTO

"main"."Tweets"("tweet_id","twitter_handle","tweet_text","created_at","retweet_count","likes_count") VALUES ("1596686265692590080","bettylo52207153","@Stickit2Stage4 Thank you so much Susan. I will let her know and its good to hear it's not unusual. I never experi... https://t.co/WJRFWOv4SG","2022-11-27 02:04:05+00:00","0","1");

Twitter_User table:

INSERT INTO

"main"."Twitter_User"("user_id","user_handle","user_name","user_profile_image","description","followers_count","created at") VALUES

("1359714993357393923","bettylo52207153","bettylou,http://abs.twimg.com/sticky/default_profile_images/default_profile_normal.png","Ocean State, D/x TNBC breast cancer", "BRCA1+ Kind always!"

,"56","2021-02-11 04:05:03+00:00");

Hospitals_sentiment_by_disease table:

INSERT INTO "main". "hospitals_sentiment_by_disease" ("Tweet Text", "sentiment", "hospital", "disease") VALUES ("Please reduce the price of perjeta as perjeta is life saving drug. Now Zydus and Intas in race of launching the par... https://t.co/SKTiQZ9Cey", "0.0", "Massachusetts General Hospital", "Alzheimers disease");

Medicines sentiment by disease table:

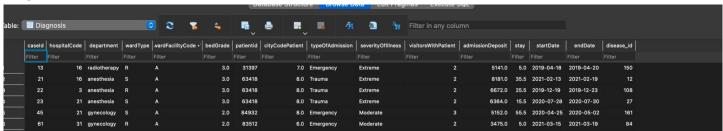
INSERT INTO "main". "medicines_sentiment_by_disease" ("Tweet Text", "sentiment", "medicine", "disease") VALUES ("Please reduce the price of perjeta as perjeta is life saving drug. Now Zydus and Intas in race of launching the par... https://t.co/SKTiQZ9Cey", "0.0", "Perjeta", "Cancer");

Query sentiment table:

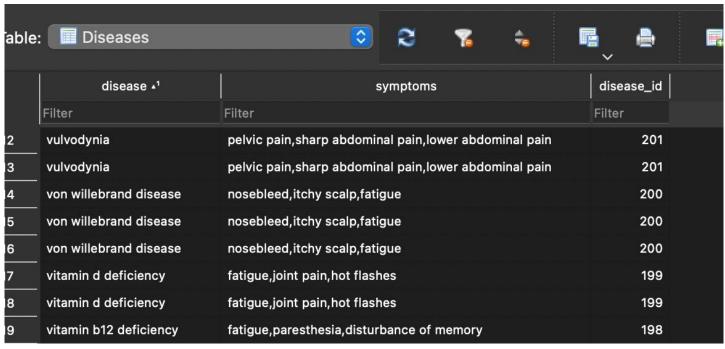
INSERT INTO "main"."query_sentiment"("tweet_id","query","sentiment","type") VALUES ("1596686265692590080","Perjeta","0.2666666666667","medicine");

Sample Data:

Diagnosis Sample Data:



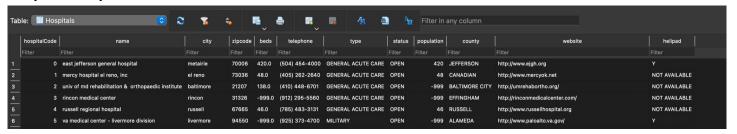
Disease Sample Data:



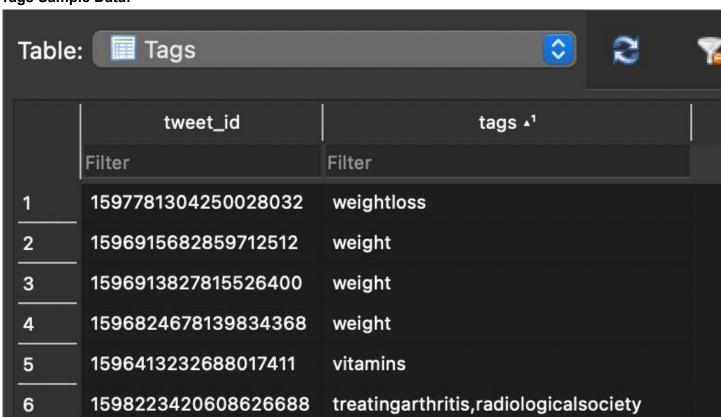
Drugs Sample Data:



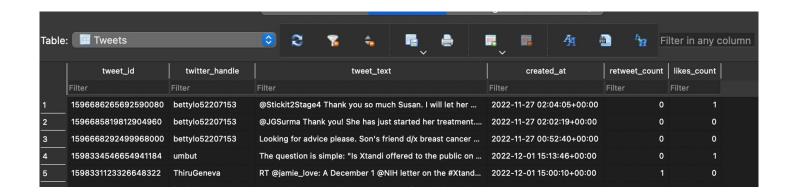
Hospital Sample Data:



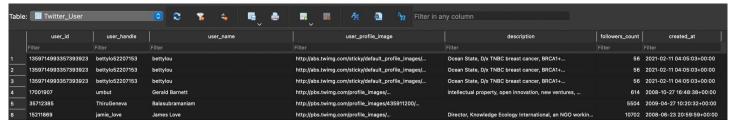
Tags Sample Data:



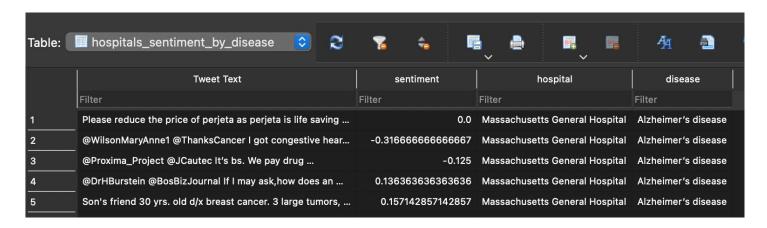
Tweets Sample Data:



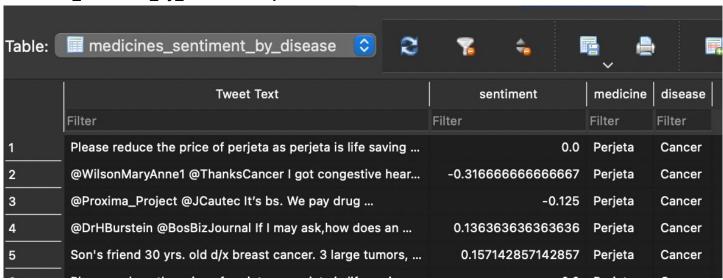
Twitter_User Sample Data:



Hospital_sentiment_by_disease Sample Data:



Medicines_sentiment_by_disease Sample Data:



query_sentiment Sample Data:

