GEMINI Programming Skills Test

Bella MacLean

May 14, 2024

1 Task One

Read in the csv files (admissions_surg.csv, admissions_med.csv and imaging.csv) and perform de-identification on all 3 files. De-identification is a common practice in health research in which an personal health identifer that identifies a patient is replaced with another unique ID for privacy purposes.

Table 1: First 5 rows of De-identified Medical Admissions

DE_ID	admission_date	admission_time	discharge_date	discharge_time	department	gender	age	main_diagnosis_icd10	main_diagnosis_name
1196	1985-10-19	04:27:00	1986-01-01	NA	General Internal Medicine	M	63	N830	Ovarian cyst
1197	1990-06-26	21:06:00	1990-09-03	21:03:00	General Internal Medicine	F	63	N410	Inflammatory conditions of male genital organs
1198	1994-06-13	06:36:00	1994-12-22	07:21:00	General Internal Medicine	F	51	K640	Hemorrhoids
1199	2005-03-11	21:54:00	2005-04-19	NA	General Internal Medicine	F	NA	C6200	Cancer of testis
1200	1997-06-26	09:49:00	1997-09-06	21:03:00	General Internal Medicine	M	81	I462	Cardiac arrest and ventricular fibrillation

Table 2: First 5 rows of De-identified Medical Admissions

DE_ID	admission_date	admission_time	discharge_date	discharge_time	department	gender	age	main_diagnosis_icd10	main_diagnosis_name
1196	1985-10-19	04:27:00	1986-01-01	NA	General Internal Medicine	M	63	N830	Ovarian cyst
1197	1990-06-26	21:06:00	1990-09-03	21:03:00	General Internal Medicine	F	63	N410	Inflammatory conditions of male genital organs
1198	1994-06-13	06:36:00	1994-12-22	07:21:00	General Internal Medicine	F	51	K640	Hemorrhoids
1199	2005-03-11	21:54:00	2005-04-19	NA	General Internal Medicine	F	NA	C6200	Cancer of testis
1200	1997-06-26	09:49:00	1997-09-06	21:03:00	General Internal Medicine	M	81	I462	Cardiac arrest and ventricular fibrillation

Table 3: First 5 rows of De-identified Imaging Data

DE_ID	test_name	ordered_date_time	performed_date	performed_time	technician_name	brief_report
1196	US	NA	1985-12-17	10:27:00	Trevon Hopson	No significant abnormality
1196	US PELVIS	NA	1985-12-02	11:40:00	Claire Melko	Indication: normal
1197	Abdomen CT	NA	1990-08-05	12:26:00	Ladonna Mcallister	Indication: Normal
1	US	NA	1998-03-24	16:15:00	claire melko	Indication: Normal
1198	CT neck + head	NA	1994-11-05	01:36:00	Lorena Burciaga	Normal

2 Task Two

Create one data frame called admissions_img, consisting of all rows in admissions_surg and admissions_med, merged with the imaging data using DE_ID (retaining all DE_IDs from both).

```
# Display the first 5 rows of the new merged dataset
head(admissions_img, 5)
```

```
  DE_ID ADMISSION.DATE ADMISSION.TIME DISCHARGE.DATE DISCHARGE.TIME
1     1     1998-02-01       07:02:00     1998-03-31       07:17:00
2     2     2010-11-30       04:33:00     2011-03-25       20:59:00
3     3     2015-03-05       09:40:00     2015-08-28       04:51:00
4     3     2015-03-05       09:40:00     2015-08-28       04:51:00
5     4     1987-11-07       21:37:00     1988-05-10       15:32:00
       DEPARTMENT GENDER AGE MAIN.DIAGNOSIS.ICD10
1 General Surgery      M  NA                E0800
2 General Surgery      F  24               M0500
3 General Surgery      M  92                O045
4 General Surgery      M  92                O045
5 General Surgery      F  93               A6000
                       MAIN.DIAGNOSIS.NAME admission_date admission_time
1     Diabetes mellitus with complications           <NA>             NA
2 Rheumatoid arthritis and related disease           <NA>             NA
3                          Induced abortion          <NA>             NA
4                          Induced abortion          <NA>             NA
5                           Viral infection          <NA>             NA
  discharge_date discharge_time department gender age main_diagnosis_icd10
1           <NA>             NA       <NA>   <NA>  NA                 <NA>
```

```
2           <NA>             NA       <NA>   <NA>  NA                 <NA>
3           <NA>             NA       <NA>   <NA>  NA                 <NA>
4           <NA>             NA       <NA>   <NA>  NA                 <NA>
5           <NA>             NA       <NA>   <NA>  NA                 <NA>
  main_diagnosis_name        test_name ordered_date_time performed_date
1                <NA>               US              <NA>     1998-03-24
2                <NA> ct neck and head              <NA>     2011-03-12
3                <NA>          ct neck              <NA>     2015-05-08
4                <NA>   RT LEG DOPPLER              <NA>     2015-05-21
5                <NA>          ct neck              <NA>     1988-01-26
  performed_time   technician_name               brief_report
1       16:15:00      claire melko         Indication: Normal
2       12:33:00   zach straughter                     Normal
3       03:06:00 mastoora al-kaber                     Cancer
4       02:44:00       marco carr On visual analysis, normal
5       11:47:00       marco carr No significant abnormality
```

3 Task Three

In admissions_img, create a new length_of_stay variable defined as discharge date and time minus admission date and time (in days). Calculate the mean length_of_stay for each department.

```
# Display the result
mean_length_of_stay_by_dept
```

```
# A tibble: 3 x 2
  DEPARTMENT      Mean_Length_of_Stay
  <chr>                         <dbl>
1 General Surgery                100.
2 Obstetrics                     105.
3 <NA>                           NaN
```

4 Task Four

In imaging, filter to the first performed test for each test_name and save the resulting data frame as q4_df. Then, transform the data into wide format such that each test_name becomes a column displaying the performed_date of that test (see example table below). Display the head of the table.

```
# Display the head of the wide format table
head(q4_df)
```

```
# A tibble: 6 x 7
     ID test_name         ordered_date_time   performed_date performed_time
  <dbl> <chr>             <dttm>              <date>         <time>
1 28711 ABDOMEN/PELVIS US NA                  1980-02-17     10:05
2 22914 Abdomen CT        NA                  1980-04-20     16:04
3 98627 CT                1980-04-16 02:26:00 1980-04-16     07:00
4 97068 CT - ABDOMEN      1980-03-14 08:58:00 1980-03-16     06:44
5 54816 CT - Femur        1981-08-14 05:30:00 1981-08-14     08:05
6 69300 CT neck + head    1980-03-24 00:59:00 1980-03-24     14:16
# i 2 more variables: technician_name <chr>, brief_report <chr>
```