Name: Amina Tabassum

NUID: 002190127

HW1 IE 7615 : Neural Networks and Deep Learning

Question 01

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```python
df=pd.read_csv('diabetic_data.csv')
df.shape
```

```
(101766, 50)
```

```python
df.dtypes
```

```
Output exceeds the size limit. Open the full output data in a text editor
encounter_id              int64
patient_nbr               int64
race                      object
gender                    object
age                       object
weight                    object
admission_type_id         int64
discharge_disposition_id  int64
admission_source_id       int64
time_in_hospital          int64
payer_code                object
medical_specialty         object
num_lab_procedures        int64
num_procedures            int64
num_medications           int64
number_outpatient         int64
number_emergency          int64
number_inpatient          int64
diag_1                    object
```
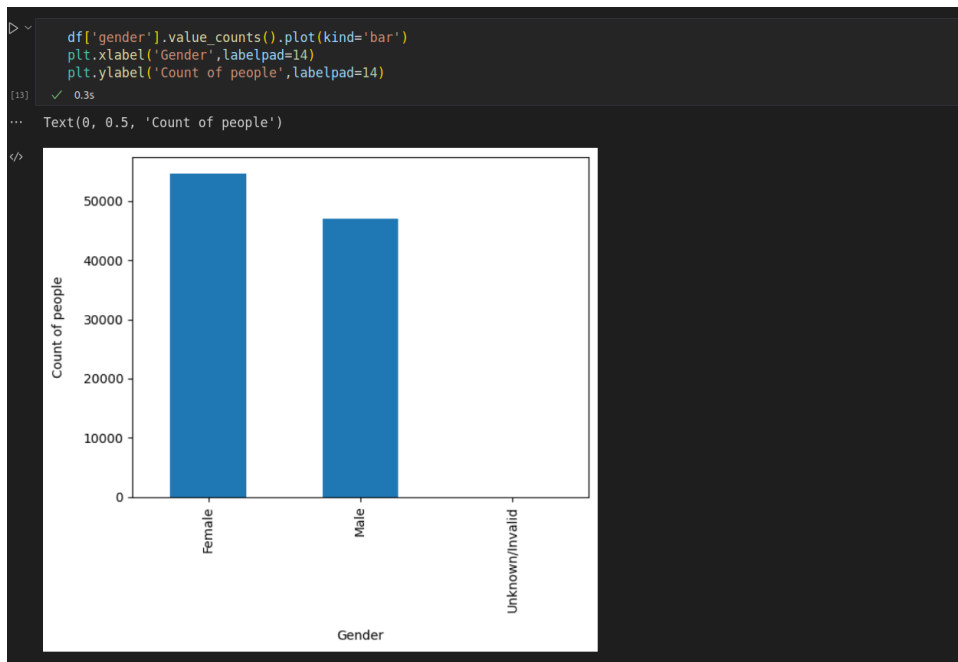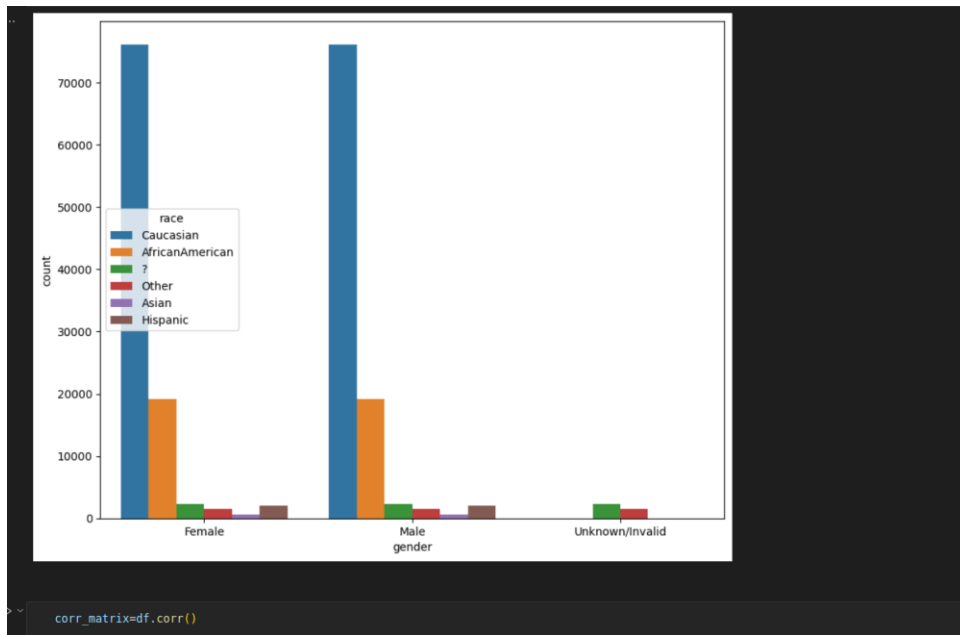
```python
df.dtypes 💡
```
[4] ✓ 0.7s

Output exceeds the size limit. Open the full output data in a text editor

```
encounter_id                  int64
patient_nbr                   int64
race                          object
gender                        object
age                           object
weight                        object
admission_type_id             int64
discharge_disposition_id      int64
admission_source_id           int64
time_in_hospital              int64
payer_code                    object
medical_specialty             object
num_lab_procedures            int64
num_procedures                int64
num_medications               int64
number_outpatient             int64
number_emergency              int64
number_inpatient              int64
diag_1                        object
diag_2                        object
diag_3                        object
number_diagnoses              int64
max_glu_serum                 object
A1Cresult                     object
metformin                     object
...
metformin-pioglitazone        object
change                        object
diabetesMed                   object
readmitted                    object
dtype: object
```

```python
df.head(10) 💡
```
[3] ✓ 0.8s   Python

	encounter_id	patient_nbr	race	gender	age	weight	admission_type_id	discharge_disposition_id	admission_source_id	time_in_hospital	...	citogliptor
0	2278392	8222157	Caucasian	Female	[0-10)	?	6	25	1	1	...	Nc
1	149190	55629189	Caucasian	Female	[10-20)	?	1	1	7	3	...	Nc
2	64410	86047875	AfricanAmerican	Female	[20-30)	?	1	1	7	2	...	Nc
3	500364	82442376	Caucasian	Male	[30-40)	?	1	1	7	2	...	Nc
4	16680	42519267	Caucasian	Male	[40-50)	?	1	1	7	1	...	Nc
5	35754	82637451	Caucasian	Male	[50-60)	?	2	1	2	3	...	Nc
6	55842	84259809	Caucasian	Male	[60-70)	?	3	1	2	4	...	Nc
7	63768	114882984	Caucasian	Male	[70-80)	?	1	1	7	5	...	Nc
8	12522	48370783	Caucasian	Female	[80-90)	?	2	1	4	13	...	Nc
9	15738	63555939	Caucasian	Female	[90-100)	?	3	3	4	12	...	Nc

0 rows × 50 columns

```python
df.describe()
```
[11] ✓ 0.1s   Python

	encounter_id	patient_nbr	admission_type_id	discharge_disposition_id	admission_source_id	time_in_hospital	num_lab_procedures	num_procedures	num_m
count	1.017660e+05	1.017660e+05	101766.000000	101766.000000	101766.000000	101766.000000	101766.000000	101766.000000	1017
mean	1.652016e+08	5.433040e+07	2.024006	3.715642	5.754437	4.395987	43.095641	1.339730	
std	1.026403e+08	3.869636e+07	1.445403	5.280166	4.064081	2.985108	19.674362	1.705807	
min	1.252200e+04	1.350000e+02	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000	
25%	8.496119e+07	2.341322e+07	1.000000	1.000000	1.000000	2.000000	31.000000	0.000000	
50%	1.523890e+08	4.550514e+07	1.000000	1.000000	7.000000	4.000000	44.000000	1.000000	
75%	2.302709e+08	8.754595e+07	3.000000	4.000000	7.000000	6.000000	57.000000	2.000000	
max	4.438672e+08	1.895026e+08	8.000000	28.000000	25.000000	14.000000	132.000000	6.000000	

```
df['gender'].value_counts().plot(kind='bar')
plt.xlabel('Gender',labelpad=14)
plt.ylabel('Count of people',labelpad=14)
```
[13]   ✓ 0.3s

...   Text(0, 0.5, 'Count of people')


```
count=df['race'].value_counts()
```
[61]   ✓ 0.5s   Python

```
df['count']=df.groupby('race')['race'].transform('count')
```
[62]   ✓ 0.5s   Python

```
fig=plt.figure(figsize=(10,8))
sns.barplot(data=df,x='gender',y='count',hue='race')
plt.show()
```
[63]   ✓ 2.1s   Python

```python
corr_matrix=df.corr()
```

corr_matrix 💡

✓ 0.7s   Python

	encounter_id	patient_nbr	admission_type_id	discharge_disposition_id	admission_source_id	time_in_hospital	num_lab_procedures	num_proce
encounter_id	1.000000	0.512028	-0.158961	-0.132876	-0.112402	-0.062221	-0.026062	-0.0
patient_nbr	0.512028	1.000000	-0.011128	-0.136814	-0.032568	-0.024092	0.015946	-0.0
admission_type_id	-0.158961	-0.011128	1.000000	0.083483	0.106654	-0.012500	-0.143713	0.1
discharge_disposition_id	-0.132876	-0.136814	0.083483	1.000000	0.018193	0.162748	0.023415	0.0
admission_source_id	-0.112402	-0.032568	0.106654	0.018193	1.000000	-0.006965	0.048885	-0.1
time_in_hospital	-0.062221	-0.024092	-0.012500	0.162748	-0.006965	1.000000	0.318450	0.1
num_lab_procedures	-0.026062	0.015946	-0.143713	0.023415	0.048885	0.318450	1.000000	0.0
num_procedures	-0.014225	-0.015570	0.129888	0.015921	-0.135400	0.191472	0.058066	1.0
num_medications	0.076113	0.020665	0.079535	0.108753	-0.054533	0.466135	0.268161	0.3
number_outpatient	0.103756	0.103379	0.026511	-0.008715	0.027244	-0.008916	-0.007602	-0.0
number_emergency	0.082803	0.062352	-0.019116	-0.024471	0.059892	-0.009681	-0.002279	-0.0
number_inpatient	0.030962	0.012480	-0.038161	0.020787	0.036314	0.073623	0.039231	-0.0
number_diagnoses	0.265149	0.226847	-0.117126	0.046891	0.072114	0.220186	0.152773	0.0
count	0.049263	0.087243	0.074117	0.015740	0.022750	-0.002518	-0.021418	0.0
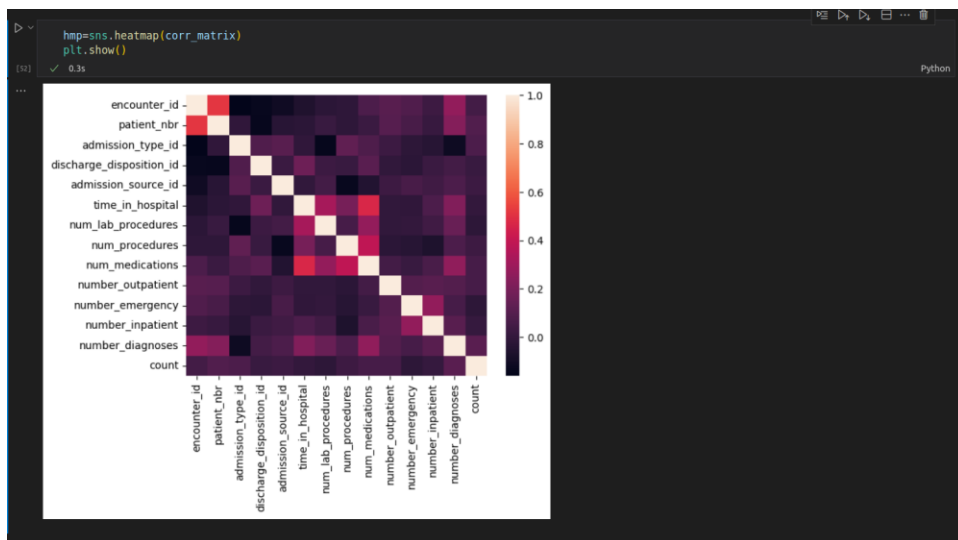
```python
hmp=sns.heatmap(corr_matrix)
plt.show()
```



```python
df.tail(15)
```

	encounter_id	patient_nbr	race	gender	age	weight	admission_type_id	discharge_disposition_id	admission_source_id	time_in_hospital	...	cit
101751	443835140	175326800	Caucasian	Male	[70-80)	?	3	6	1	13	...	
101752	443835512	139605341	Other	Female	[40-50)	?	3	1	1	3	...	
101753	443841992	184875899	Other	Male	[40-50)	?	1	1	7	13	...	
101754	443842016	183087545	Caucasian	Female	[70-80)	?	1	1	7	9	...	
101755	443842022	188574944	Other	Female	[40-50)	?	1	1	7	14	...	
101756	443842070	140199494	Other	Female	[60-70)	?	1	1	7	2	...	
101757	443842136	181593374	Caucasian	Female	[70-80)	?	1	1	7	5	...	
101758	443842340	120975314	Caucasian	Female	[80-90)	?	1	1	7	5	...	
101759	443842778	86472243	Caucasian	Male	[80-90)	?	1	1	7	1	...	
101760	443847176	50375628	AfricanAmerican	Female	[60-70)	?	1	1	7	6	...	
101761	443847548	100162476	AfricanAmerican	Male	[70-80)	?	1	3	7	3	...	
101762	443847782	74694222	AfricanAmerican	Female	[80-90)	?	1	4	5	5	...	
101763	443854148	41088789	Caucasian	Male	[70-80)	?	1	1	7	1	...	
101764	443857166	31693671	Caucasian	Female	[80-90)	?	2	3	7	10	...	
101765	443867222	175429310	Caucasian	Male	[70-80)	?	1	1	7	6	...	

15 rows × 50 columns
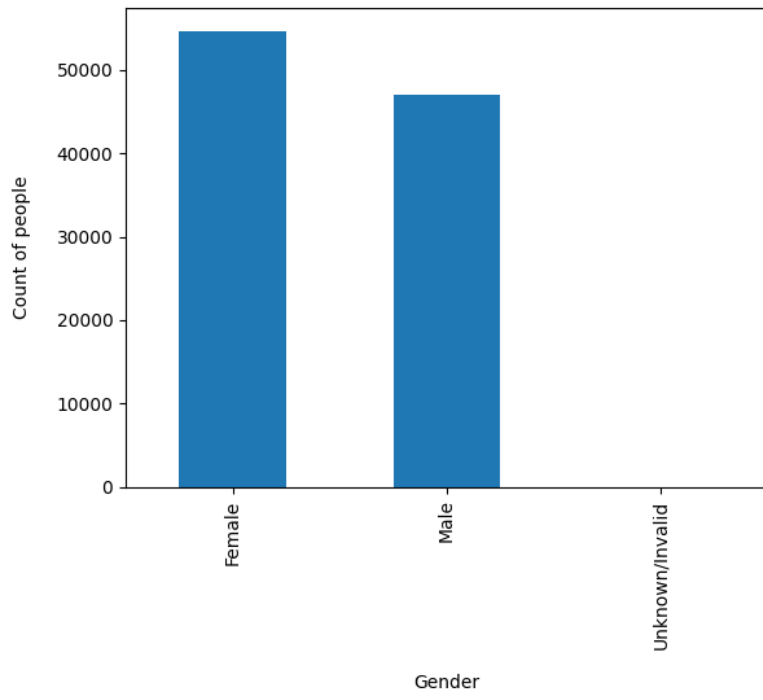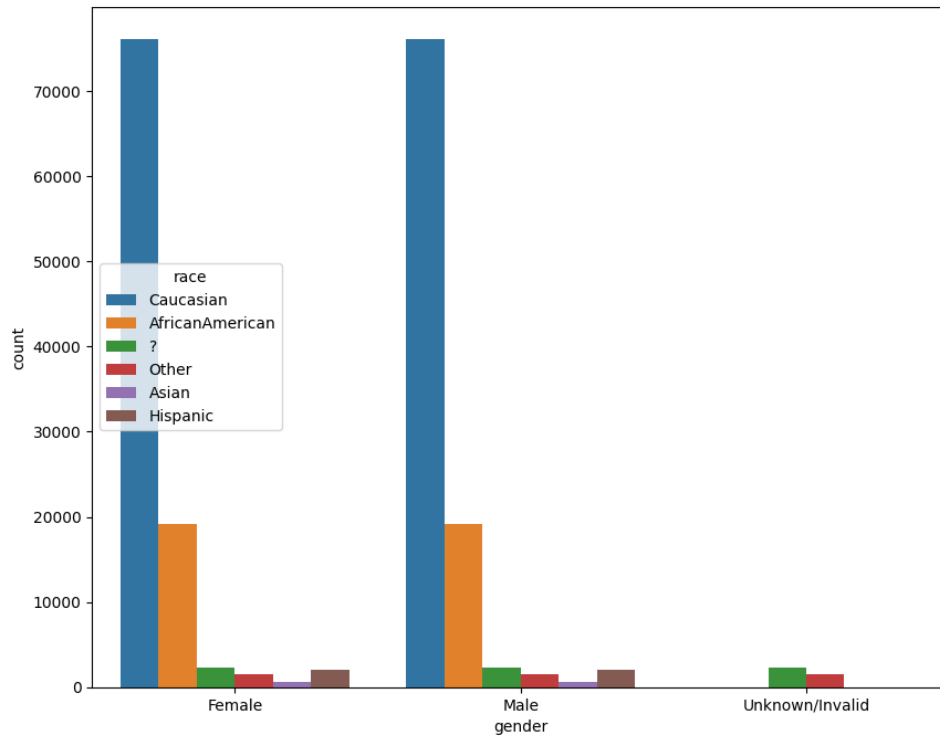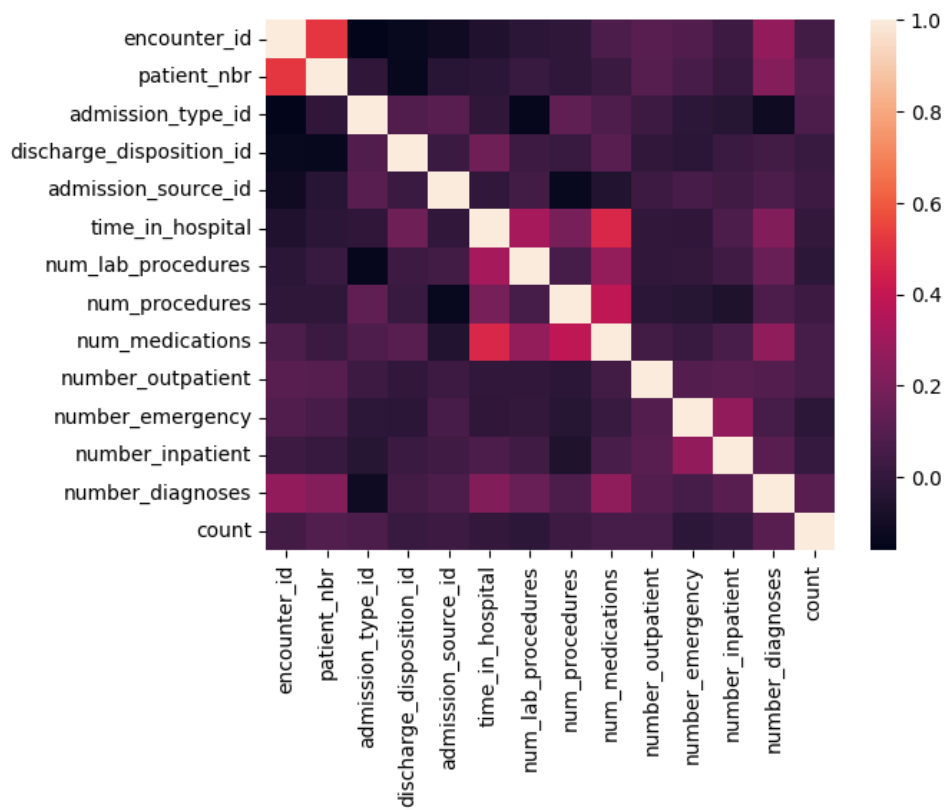
Code:

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df=pd.read_csv('diabetic_data.csv')
df.shape

df.dtypes

df.head(10)
df.tail(15)

df.describe()

df['gender'].value_counts().plot(kind='bar')
plt.xlabel('Gender',labelpad=14)
plt.ylabel('Count of people',labelpad=14)

count=df['race'].value_counts()

df['count']=df.groupby('race')['race'].transform('count')

fig=plt.figure(figsize=(10,8))
sns.barplot(data=df,x='gender',y='count',hue='race')
plt.show()
corr_matrix=df.corr()
corr_matrix
hmp=sns.heatmap(corr_matrix)
plt.show()
```