In [2]:
```
!python --version
```

```
Python 3.7.10
```

In [3]:
```
!pip install --disable-pip-version-check -q sagemaker==2.38.0
!pip install --disable-pip-version-check -q smdebug==1.0.4
!pip install --disable-pip-version-check -q sagemaker-experiments==0.1.28
```

```
/opt/conda/lib/python3.7/site-packages/secretstorage/dhcrypto.py:16: Cryptograp
hyDeprecationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
/opt/conda/lib/python3.7/site-packages/secretstorage/util.py:25: CryptographyDe
precationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
WARNING: Running pip as the 'root' user can result in broken permissions and co
nflicting behaviour with the system package manager. It is recommended to use a
virtual environment instead: https://pip.pypa.io/warnings/venv
/opt/conda/lib/python3.7/site-packages/secretstorage/dhcrypto.py:16: Cryptograp
hyDeprecationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
/opt/conda/lib/python3.7/site-packages/secretstorage/util.py:25: CryptographyDe
precationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
WARNING: Running pip as the 'root' user can result in broken permissions and co
nflicting behaviour with the system package manager. It is recommended to use a
virtual environment instead: https://pip.pypa.io/warnings/venv
/opt/conda/lib/python3.7/site-packages/secretstorage/dhcrypto.py:16: Cryptograp
hyDeprecationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
/opt/conda/lib/python3.7/site-packages/secretstorage/util.py:25: CryptographyDe
precationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
WARNING: Running pip as the 'root' user can result in broken permissions and co
nflicting behaviour with the system package manager. It is recommended to use a
virtual environment instead: https://pip.pypa.io/warnings/venv
```

In [4]:
```
!pip install --disable-pip-version-check -q tensorflow==2.3.1
```

```
/opt/conda/lib/python3.7/site-packages/secretstorage/dhcrypto.py:16: Cryptograp
hyDeprecationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
/opt/conda/lib/python3.7/site-packages/secretstorage/util.py:25: CryptographyDe
precationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
WARNING: Running pip as the 'root' user can result in broken permissions and co
nflicting behaviour with the system package manager. It is recommended to use a
virtual environment instead: https://pip.pypa.io/warnings/venv
```

In [5]:
```
!pip install --disable-pip-version-check -q tensorflow==2.3.1
```

```
/opt/conda/lib/python3.7/site-packages/secretstorage/dhcrypto.py:16: Cryptog
raphyDeprecationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
/opt/conda/lib/python3.7/site-packages/secretstorage/util.py:25: CryptographyDe
precationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
WARNING: Running pip as the 'root' user can result in broken permissions and co
nflicting behaviour with the system package manager. It is recommended to use a
virtual environment instead: https://pip.pypa.io/warnings/venv
```

In [6]:
```
!pip install --disable-pip-version-check -q transformers==3.5.1
```

```
/opt/conda/lib/python3.7/site-packages/secretstorage/dhcrypto.py:16: Cryptog
raphyDeprecationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
/opt/conda/lib/python3.7/site-packages/secretstorage/util.py:25: CryptographyDe
precationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
WARNING: Running pip as the 'root' user can result in broken permissions and co
nflicting behaviour with the system package manager. It is recommended to use a
virtual environment instead: https://pip.pypa.io/warnings/venv
```

In [7]:
```
!pip install --disable-pip-version-check -q PyAthena==2.1.1
```

```
/opt/conda/lib/python3.7/site-packages/secretstorage/dhcrypto.py:16: Cryptog
raphyDeprecationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
/opt/conda/lib/python3.7/site-packages/secretstorage/util.py:25: CryptographyDe
precationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
WARNING: Running pip as the 'root' user can result in broken permissions and co
nflicting behaviour with the system package manager. It is recommended to use a
virtual environment instead: https://pip.pypa.io/warnings/venv
```

In [8]:
```
!pip install --disable-pip-version-check -q SQLAlchemy==1.3.23
```

```
/opt/conda/lib/python3.7/site-packages/secretstorage/dhcrypto.py:16: Cryptog
raphyDeprecationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
/opt/conda/lib/python3.7/site-packages/secretstorage/util.py:25: CryptographyDe
precationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
WARNING: Running pip as the 'root' user can result in broken permissions and co
nflicting behaviour with the system package manager. It is recommended to use a
virtual environment instead: https://pip.pypa.io/warnings/venv
```

In [9]:
```
!conda install -q -y zip
```

```
Collecting package metadata (current_repodata.json): ...working... done
Solving environment: ...working... done

# All requested packages already installed.
```

In [10]:

```
!pip install --disable-pip-version-check -q matplotlib==3.1.3
```

```
/opt/conda/lib/python3.7/site-packages/secretstorage/dhcrypto.py:16: Cryptograp
hyDeprecationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
/opt/conda/lib/python3.7/site-packages/secretstorage/util.py:25: CryptographyDe
precationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
WARNING: Running pip as the 'root' user can result in broken permissions and co
nflicting behaviour with the system package manager. It is recommended to use a
virtual environment instead: https://pip.pypa.io/warnings/venv
```

In [11]:

```
!pip install --disable-pip-version-check -q seaborn==0.10.0
```

```
/opt/conda/lib/python3.7/site-packages/secretstorage/dhcrypto.py:16: Cryptograp
hyDeprecationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
/opt/conda/lib/python3.7/site-packages/secretstorage/util.py:25: CryptographyDe
precationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
WARNING: Running pip as the 'root' user can result in broken permissions and co
nflicting behaviour with the system package manager. It is recommended to use a
virtual environment instead: https://pip.pypa.io/warnings/venv
```

In [12]:

```
!pip list
```

```
/opt/conda/lib/python3.7/site-packages/secretstorage/dhcrypto.py:16: Cryptograp
hyDeprecationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
/opt/conda/lib/python3.7/site-packages/secretstorage/util.py:25: CryptographyDe
precationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
Package                              Version
------------------------------------ -----------------
absl-py                              1.0.0
aiobotocore                          2.0.1
aiohttp                              3.8.1
aioitertools                         0.8.0
aiosignal                            1.2.0
alabaster                            0.7.12
anaconda-client                      1.7.2
anaconda-project                     0.8.3
argh                                 0.26.2
argon2-cffi                          21.3.0
argon2-cffi-bindings                 21.2.0
asn1crypto                           1.3.0
astroid                              2.9.0
astropy                              4.0
astunparse                           1.6.3
async-timeout                        4.0.1
asynctest                            0.13.0
atomicwrites                         1.3.0
attrs                                19.3.0
autopep8                             1.4.4
autovizwidget                        0.19.1
awscli                               1.18.216
awswrangler                          2.3.0
Babel                                2.9.1
backcall                             0.1.0
backports.shutil-get-terminal-size   1.0.0
beautifulsoup4                       4.8.2
bitarray                             1.2.1
bkcharts                             0.2
bleach                               4.1.0
bokeh                                1.4.0
boto                                 2.49.0
boto3                                1.16.56
botocore                             1.19.56
Bottleneck                           1.3.2
brotlipy                             0.7.0
cached-property                      1.5.2
cachetools                           5.0.0
certifi                              2021.10.8
cffi                                 1.14.6
chardet                              3.0.4
charset-normalizer                   2.0.4
Click                                7.0
cloudpickle                          2.0.0
clyent                               1.2.2
colorama                             0.4.3
conda                                4.12.0
conda-package-handling               1.7.3
contextlib2                          0.6.0.post1
cryptography                         36.0.0
cycler                               0.10.0
Cython                               0.29.15
cytoolz                              0.10.1
```

```
dask                                 2021.12.0
decorator                            4.4.1
defusedxml                           0.6.0
diff-match-patch                     20181111
dill                                 0.3.4
distributed                          2021.12.0
distro                               1.6.0
docker                               5.0.0
docker-compose                       1.29.2
dockerpty                            0.4.1
docopt                               0.6.2
docutils                             0.15.2
dparse                               0.5.1
entrypoints                          0.3
enum-compat                          0.0.3
et-xmlfile                           1.0.1
fastcache                            1.1.0
filelock                             3.0.12
flake8                               3.7.9
Flask                                1.1.1
frozenlist                           1.2.0
fsspec                               2021.11.1
future                               0.18.2
gast                                 0.3.3
gevent                               1.4.0
glob2                                0.7
gmpy2                                2.0.8
google-auth                          2.6.2
google-auth-oauthlib                 0.4.6
google-pasta                         0.2.0
greenlet                             0.4.15
grpcio                               1.45.0
h5py                                 2.10.0
hdijupyterutils                      0.19.1
HeapDict                             1.0.1
html5lib                             1.0.1
hypothesis                           5.5.4
idna                                 2.8
imageio                              2.6.1
imagesize                            1.2.0
importlib-metadata                   4.11.3
intervaltree                         3.0.2
ipykernel                            5.1.4
ipython                              7.12.0
ipython_genutils                     0.2.0
ipywidgets                           7.5.1
isort                                4.3.21
itsdangerous                         1.1.0
jdcal                                1.4.1
jedi                                 0.14.1
jeepney                              0.4.2
Jinja2                               3.0.3
jmespath                             0.10.0
joblib                               0.14.1
json5                                0.9.1
jsonschema                           3.2.0
jupyter                              1.0.0
jupyter-client                       5.3.4
jupyter-console                      6.1.0
jupyter-core                         4.6.1
jupyterlab                           1.2.21
jupyterlab-server                    1.0.6
```

```
Keras-Preprocessing                  1.1.2
keyring                              21.1.0
kiwisolver                           1.1.0
lazy-object-proxy                    1.4.3
libarchive-c                         2.8
lief                                 0.9.0
llvmlite                             0.37.0
locket                               0.2.0
lxml                                 4.8.0
Markdown                             3.3.6
MarkupSafe                           2.0.1
matplotlib                           3.1.3
mccabe                               0.6.1
mistune                              0.8.4
mkl-fft                              1.0.15
mkl-random                           1.1.0
mkl-service                          2.3.0
mock                                 4.0.1
more-itertools                       8.2.0
mpmath                               1.1.0
msgpack                              0.6.1
multidict                            5.2.0
multipledispatch                     0.6.0
multiprocess                         0.70.12.2
nbconvert                            5.6.1
nbformat                             5.0.4
nest-asyncio                         1.5.4
networkx                             2.4
nltk                                 3.4.5
nose                                 1.3.7
notebook                             6.4.6
numba                                0.54.1
numexpr                              2.7.1
numpy                                1.18.5
numpydoc                             0.9.2
oauthlib                             3.2.0
olefile                              0.46
openpyxl                             3.0.3
opt-einsum                           3.3.0
packaging                            20.1
pandas                               1.2.0
pandocfilters                        1.4.2
parso                                0.5.2
partd                                1.1.0
path                                 13.1.0
pathlib2                             2.3.5
pathos                               0.2.8
pathtools                            0.1.2
patsy                                0.5.1
pep8                                 1.7.1
pexpect                              4.8.0
pg8000                               1.16.6
pickleshare                          0.7.5
Pillow                               8.4.0
pip                                  22.0.4
pkginfo                              1.5.0.1
platformdirs                         2.4.0
plotly                               5.4.0
pluggy                               0.13.1
ply                                  3.11
pox                                  0.3.0
ppft                                 1.6.6.4
```

```
prometheus-client                    0.7.1
prompt-toolkit                       3.0.3
protobuf                             3.19.1
protobuf3-to-dict                    0.1.5
psutil                               5.6.7
ptyprocess                           0.6.0
pure-sasl                            0.6.2
py                                   1.11.0
pyarrow                              2.0.0
pyasn1                               0.4.8
pyasn1-modules                       0.2.8
pyathena                             2.1.1
pycodestyle                          2.5.0
pycosat                              0.6.3
pycparser                            2.19
pycrypto                             2.6.1
pycurl                               7.43.0.5
pydocstyle                           4.0.1
pyflakes                             2.1.1
pyfunctional                         1.4.3
Pygments                             2.5.2
PyHive                               0.6.4
pyinstrument                         4.1.1
pykerberos                           1.2.1
pylint                               2.12.2
PyMySQL                              1.0.2
pyodbc                               4.0.0-unsupported
pyOpenSSL                            19.1.0
pyparsing                            2.4.6
pyrsistent                           0.15.7
PySocks                              1.7.1
pytest                               5.3.5
pytest-arraydiff                     0.3
pytest-astropy                       0.8.0
pytest-astropy-header                0.1.2
pytest-doctestplus                   0.5.0
pytest-openfiles                     0.4.0
pytest-remotedata                    0.3.2
python-dateutil                      2.8.1
python-dotenv                        0.19.2
python-jsonrpc-server                0.3.4
python-language-server               0.31.7
pytz                                 2021.3
PyWavelets                           1.1.1
pyxdg                                0.26
PyYAML                               5.3.1
pyzmq                                18.1.1
QDarkStyle                           2.8
QtAwesome                            0.6.1
qtconsole                            4.6.0
QtPy                                 1.9.0
redshift-connector                   2.0.905
regex                                2022.3.15
requests                             2.26.0
requests-kerberos                    0.12.0
requests-oauthlib                    1.3.1
rope                                 0.16.0
rsa                                  4.5
Rtree                                0.9.3
ruamel_yaml                          0.15.87
s3fs                                 2021.11.1
s3transfer                           0.3.7
```

```
sacremoses                           0.0.49
sagemaker                            2.38.0
sagemaker-experiments                0.1.28
sagemaker-studio-analytics-extension 0.0.4
sagemaker-studio-sparkmagic-lib      0.1.3
sasl                                 0.2.1
scikit-image                         0.16.2
scikit-learn                         0.22.1
scipy                                1.4.1
scramp                               1.2.0
seaborn                              0.10.0
SecretStorage                        3.1.2
Send2Trash                           1.8.0
sentencepiece                        0.1.91
setuptools                           59.5.0
simplegeneric                        0.8.1
singledispatch                       3.4.0.3
six                                  1.14.0
sklearn                              0.0
smclarify                            0.2
smdebug                              1.0.4
smdebug-rulesconfig                  1.0.1
snowballstemmer                      2.0.0
sortedcollections                    1.1.2
sortedcontainers                     2.1.0
soupsieve                            1.9.5
sparkmagic                           0.19.1
Sphinx                               2.4.0
sphinxcontrib-applehelp              1.0.1
sphinxcontrib-devhelp                1.0.1
sphinxcontrib-htmlhelp               1.0.2
sphinxcontrib-jsmath                 1.0.1
sphinxcontrib-qthelp                 1.0.2
sphinxcontrib-serializinghtml        1.1.3
sphinxcontrib-websupport             1.2.0
spyder                               4.0.1
spyder-kernels                       1.8.1
SQLAlchemy                           1.3.23
statsmodels                          0.11.0
stepfunctions                        2.0.0rc1
sympy                                1.5.1
tables                               3.6.1
tabulate                             0.8.9
tblib                                1.6.0
tenacity                             8.0.1
tensorboard                          2.8.0
tensorboard-data-server              0.6.1
tensorboard-plugin-wit               1.8.1
tensorflow                           2.3.1
tensorflow-estimator                 2.3.0
termcolor                            1.1.0
terminado                            0.8.3
testpath                             0.4.4
texttable                            1.6.4
thrift                               0.13.0
thrift-sasl                          0.4.3
tokenizers                           0.9.3
toml                                 0.10.2
toolz                                0.10.0
torch                                1.6.0
torch-model-archiver                 0.3.0
torchserve                           0.3.0
```

```
tornado                              6.1
tqdm                                 4.42.1
traitlets                            4.3.3
transformers                         3.5.1
typed-ast                            1.5.1
typing_extensions                    4.0.1
ujson                                1.35
unicodecsv                           0.14.1
urllib3                              1.26.7
watchdog                             0.10.2
wcwidth                              0.1.8
webencodings                         0.5.1
websocket-client                     0.59.0
Werkzeug                             1.0.0
wheel                                0.34.2
widgetsnbextension                   3.5.1
wrapt                                1.14.0
wurlitzer                            2.0.0
xlrd                                 1.2.0
XlsxWriter                           1.2.7
xlwt                                 1.3.0
yapf                                 0.28.0
yarl                                 1.7.2
zict                                 1.0.0
zipp                                 2.2.0
```

In [13]:
```
setup_dependencies_passed = True
```

In [14]:
```
%store
```

```
Stored variables and their in-db values:
autopilot_train_s3_uri                                -> 's3://sagemaker-us-eas
t-1-189468192453/data/amazon
ingest_create_athena_table_parquet_passed             -> True
s3_private_path_tsv                                   -> 's3://sagemaker-us-eas
t-1-189468192453/ads-508-azh
s3_public_path_tsv                                    -> 's3://ads-508-azhang/f
inalproject/'
setup_dependencies_passed                             -> True
setup_iam_roles_passed                                -> True
setup_instance_check_passed                           -> True
setup_s3_bucket_passed                                -> True
```

In [15]:
```
from pyathena import connect
import pandas as pd
```

In [16]:
```
%store -r setup_dependencies_passed
```

In [17]:
```
try:
    setup_dependencies_passed
except NameError:
    print("++++++++++++++++++++++++++++++++++++++++++++++")
    print("[ERROR] YOU HAVE TO RUN THE PREVIOUS NOTEBOOK ")
    print("You did not install the required libraries.   ")
    print("++++++++++++++++++++++++++++++++++++++++++++++")
```

In [18]:

```
print(setup_dependencies_passed)
```

```
True
```

In [19]:

```
if not setup_dependencies_passed:
    print("++++++++++++++++++++++++++++++++++++++++++++++")
    print("[ERROR] YOU HAVE TO RUN THE PREVIOUS NOTEBOOK ")
    print("You did not install the required libraries.   ")
    print("++++++++++++++++++++++++++++++++++++++++++++++")
else:
    print("[OK]")
```

```
[OK]
```

In [20]:

```
%store -r setup_iam_roles_passed
```

In [21]:

```
try:
    setup_iam_roles_passed
except NameError:
    print("+++++++++++++++++++++++++++++++")
    print("[ERROR] YOU HAVE TO RUN ALL NOTEBOOKS IN THE SETUP FOLDER FIRST. You
are missing Setup IAM Roles.")
    print("+++++++++++++++++++++++++++++++")
```

In [22]:

```
print(setup_iam_roles_passed)
```

```
True
```

In [23]:

```
import boto3

region = boto3.Session().region_name
session = boto3.session.Session()

ec2 = boto3.Session().client(service_name="ec2", region_name=region)
sm = boto3.Session().client(service_name="sagemaker", region_name=region)
```

In [24]:

```
import json

notebook_instance_name = None

try:
    with open("/opt/ml/metadata/resource-metadata.json") as notebook_info:
        data = json.load(notebook_info)
        domain_id = data["DomainId"]
        resource_arn = data["ResourceArn"]
        region = resource_arn.split(":")[3]
        name = data["ResourceName"]
    print("DomainId: {}".format(domain_id))
    print("Name: {}".format(name))
except:
    print("+++++++++++++++++++++++++++++++++++++++++")
    print("[ERROR]: COULD NOT RETRIEVE THE METADATA.")
    print("+++++++++++++++++++++++++++++++++++++++++")
```

```
DomainId: d-2ywrjjiz4zpt
Name: datascience-1-0-ml-t3-medium-1abf3407f667f989be9d86559395
```

In [25]:

```
%store -r setup_instance_check_passed
```

In [26]:
```
try:
    setup_instance_check_passed
except NameError:
    print("++++++++++++++++++++++++++++++++")
    print("[ERROR] YOU HAVE TO RUN ALL NOTEBOOKS IN THE SETUP FOLDER FIRST. You are missing Instance Check.")
    print("++++++++++++++++++++++++++++++++")
```

In [27]:
```
%store -r setup_dependencies_passed
```

In [28]:
```
try:
    setup_dependencies_passed
except NameError:
    print("++++++++++++++++++++++++++++++++")
    print("[ERROR] YOU HAVE TO RUN ALL NOTEBOOKS IN THE SETUP FOLDER FIRST. You are missing Setup Dependencies.")
    print("++++++++++++++++++++++++++++++++")
```

In [29]:
```
print(setup_dependencies_passed)
```

```
True
```

In [30]:
```
%store -r setup_s3_bucket_passed
```

In [31]:
```
try:
    setup_s3_bucket_passed
except NameError:
    print("++++++++++++++++++++++++++++++++")
    print("[ERROR] YOU HAVE TO RUN ALL NOTEBOOKS IN THE SETUP FOLDER FIRST. You are missing Setup Dependencies.")
    print("++++++++++++++++++++++++++++++++")
```

In [32]:
```
print(setup_s3_bucket_passed)
```

```
True
```

In [33]:
```
if not setup_instance_check_passed:
    print("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
    print("[ERROR] YOU HAVE TO RUN ALL NOTEBOOKS IN THE SETUP FOLDER FIRST. You are missing Instance Check.")
    print("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
if not setup_dependencies_passed:
    print("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
    print("[ERROR] YOU HAVE TO RUN ALL NOTEBOOKS IN THE SETUP FOLDER FIRST. You are missing Setup Dependencies.")
    print("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
if not setup_s3_bucket_passed:
    print("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
    print("[ERROR] YOU HAVE TO RUN ALL NOTEBOOKS IN THE SETUP FOLDER FIRST. You are missing Setup S3 Bucket.")
    print("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
if not setup_iam_roles_passed:
    print("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
    print("[ERROR] YOU HAVE TO RUN ALL NOTEBOOKS IN THE SETUP FOLDER FIRST. You are missing Setup IAM Roles.")
    print("+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++")
```

In [34]:

```
!aws s3 ls s3://ads-508-azhang/finalproject/
```

```
2022-03-21 23:41:03          0 
2022-03-22 02:21:16   11743774 NHSDA-1979-DS0001-data-excel.tsv
2022-03-21 23:41:22   22765996 NHSDA-1988-DS0001-data-excel.tsv
2022-03-21 23:41:22   58394554 NHSDA-1995-DS0001-data-excel.tsv
```

In [35]:

```
import boto3
import sagemaker
import pandas as pd

sess = sagemaker.Session()
bucket = sess.default_bucket()
role = sagemaker.get_execution_role()
region = boto3.Session().region_name
account_id = boto3.client("sts").get_caller_identity().get("Account")

sm = boto3.Session().client(service_name="sagemaker", region_name=region)
```

In [36]:

```
s3_public_path_tsv = "s3://ads-508-azhang/finalproject/"
```

In [37]:

```
%store s3_public_path_tsv
```

```
Stored 's3_public_path_tsv' (str)
```

In [38]:

```
s3_private_path_tsv = "s3://{}/ads-508-azhang/finalproject".format(bucket)
```

In [39]:

```
print(s3_private_path_tsv)
```

```
s3://sagemaker-us-east-1-189468192453/ads-508-azhang/finalproject
```

In [44]:

```
%store s3_private_path_tsv
```

```
Stored 's3_private_path_tsv' (str)
```

In [52]:

```
!aws s3 cp --recursive $s3_public_path_tsv/ $s3_private_path_tsv/ --exclude "*" --include "NHSDA-1988-DS0001-data-excel.tsv"
!aws s3 cp --recursive $s3_public_path_tsv/ $s3_private_path_tsv/ --exclude "*" --include "NHSDA-1995-DS0001-data-excel.tsv"
!aws s3 cp --recursive $s3_public_path_tsv/ $s3_private_path_tsv/ --exclude "*" --include "NHSDA-1979-DS0001-data-excel.tsv"
```

In [46]:

```
print(s3_private_path_tsv)
```

```
s3://sagemaker-us-east-1-189468192453/ads-508-azhang/finalproject
```

In [60]:

```
!aws s3 ls $s3_private_path_tsv/
```

```
2022-03-24 02:19:54   11743774 NHSDA-1979-DS0001-data-excel.tsv
2022-03-24 02:19:03   22765996 NHSDA-1988-DS0001-data-excel.tsv
2022-03-24 02:19:03   58394554 NHSDA-1995-DS0001-data-excel.tsv
```

In [48]:

```
session = boto3.Session()


#Then use the session to get the resource
s3 = session.resource('s3')

my_bucket = s3.Bucket('ads-508-azhang')

for my_bucket_object in my_bucket.objects.all():
    print(my_bucket_object.key)
```

```
finalproject/
finalproject/NHSDA-1979-DS0001-data-excel.tsv
finalproject/NHSDA-1988-DS0001-data-excel.tsv
finalproject/NHSDA-1995-DS0001-data-excel.tsv
```

In [61]:

```
from IPython.core.display import display, HTML

display(
    HTML(
        '<b>Review <a target="blank" href="https://s3.console.aws.amazon.com/s
3/buckets/sagemaker-{}-{}/ads-508-azhang/finalproject/?region={}&tab=overview">
S3 Bucket</a></b>'.format(
            region, account_id, region
        )
    )
)
```

Review S3 Bucket (https://s3.console.aws.amazon.com/s3/buckets/sagemaker-us-east-1-189468192453/ads-508-azhang/finalproject/?region=us-east-1&tab=overview)

In [50]:

```
%store
```

```
Stored variables and their in-db values:
autopilot_train_s3_uri                                 -> 's3://sagemaker-us-eas
t-1-189468192453/data/amazon
ingest_create_athena_table_parquet_passed              -> True
s3_private_path_tsv                                    -> 's3://sagemaker-us-eas
t-1-189468192453/ads-508-azh
s3_public_path_tsv                                     -> 's3://ads-508-azhang/f
inalproject/'
setup_dependencies_passed                              -> True
setup_iam_roles_passed                                 -> True
setup_instance_check_passed                            -> True
setup_s3_bucket_passed                                 -> True
```

In [53]:

```
!aws s3 cp s3://ads-508-azhang/finalproject/NHSDA-1979-DS0001-data-excel.tsv -
| head
```

```
CASEID  RESPID  ENCPSU  ENCSEG  ENCCASE CIGMORLS        CIGTRY  CIG5PK  CIGREC
AVCIG   HRDHER  HRDMJ   HRDCOC  HRDLSD  HRDBAR  HRDTRN  HRDAMP  ADDHER  ADDALC
ADDMJ   ADDTOB  ADDBAR  ADDTRN  ADDAMP  ADDLSD  ADDCOC  ADDNONE SEDLIKE SEDFEEL
SEDNEED SEDREC  SED30MOA        SED30MOB        SED30MOC        SEDDAL30
BUTISOL BUTICAPS        AMYTAL  ESKABARB        LUMINAL MEBARAL AMOBARB PHENOBA
R       ALURATE PLACIDYL        DORIDEN NOLUDAR SOPOR   QUAALUDE        PAREST
NOCTEC  METHAQ  CHHYD   NEMBUTAL        CARBTAL SECONAL TUINAL  PENTOB  SECOB
DALMANE SEDDKNAM        NOSEDAT SEDAGE  TRNLIKE TRNFEEL TRNNEED TRANREC TRN30MO
A       TRN30MOB        TRN30MOC        TRNBEN30        VALIUM  LIBRIUM LIBRITA
B       SKLY    SERAX   TRANXENE        ATIVAN  VERSTRAN        MEPRSPAN
MILTOWN EQUANIL MEPROB  VISTAR  ATARAX  BENADRYL        TRDKNAM NOTRANQ TRANAGE
STIMLIKE        STIMFEEL        STIMNEED        STIMREC STM30MOA        STM30MO
B       STMRIT30        STMCYL30        DEXED   DEXAMYL ESKAT   BENZ    BIPHET
DESOXYN DETAMP  METHI   OBLA    TENUATE TEPANIL DIDREX  PLEGINE PRELUDIN
PRESATE IONAMIN PONDIMIN        VORANIL SANOREX RITALIN CYLERT  STMDKNAM
NOSTIMS STIMAGE ANALLIKE        ANALFEEL        ANALNEED        ANALREC ANL30MO
A       ANL30MOB        ANL30MOC        ANLTAL30        DARVON  DOLENE  SK65A
PROPOXY LERITINE        LEVODRO PERCODAN        DEMEROL DILAUD  TYLCOD  CODEINE
DOLOP   WESTODON        METHDON TALWIN  ANLDKNAM        ANALNONE        ANALAGE
ALCFIRST        ALCTRY  ALCREC  ALCDAYS MODR30A MODR30DY        UNDSTAS1
VRA7AS1 MRKEAAS1        VRA8AS1 MJKNOWN MJOPP   MJFIRST MJAGE   MJLIVE  MJREC
MJDAY30A        MJTOT   UNDSTAS2        VRM9AS2 MRKEAAS2        VRM10AS2
INHREAD INHOPP  INHFIRST        INHAGE  GAS     SPPAINT AEROS   GLUE    SOLVENT
AMYLNIT ETHER   NITOXID ODORIZER        INHNEVER        GAS30A  SPPAN30A
AEROS30A        GLUE30A SOLVN30A        AMLNT30A        ETHER30A        NOX30A
ODR30A  INH30NO INHREC  INHTOT  INHODRHR        INHODRUS        UNDSTAS3
VRG10AS3        MRKEAAS3        VRG11AS3        HALLOPP HALFIRST        HALLAGE
HALLREC HAL30USE        HALLTOT HALPCPHR        PCP     HALPCP30        UNDSTAS
4       VRL10AS4        MRKEAAS4        VRL11AS4        COCOPP  COCFIRST
COCAGE  COCREC  COCUS30A        COCTOT  UNDSTAS5        VRC7AS5 MRKEAAS5
VRC8AS5 HERKNOW HEROPP  HERFIRST        HERAGE  HERREC  HER30USE        HERTOT
HERFRNDS        HERNOADR        HERNEEDL        UNDSTAS6        VRH11AS6
MRKEAAS6        VRH12AS6        SPLCOC  SPLHAL  SPLCIG  SPLHER  SPLBEER SPLLQR
SPLMJR  SPLPILLS        SPLINH  GMJNOHO GMJNONE GMJMED  GMJJOB  GMJFUN  GMJRELA
X       GMJAWARE        GMJCNFDN        GMJDEAL GMJSLEEP        GMJSEX  GMJAPPE
T       GMJDK   GMJMISC GMJREF1 BMJCONTR        BMJMEMRY        BMJNONE BMJHABI
T       BMJSTRGR        BMJHLTH BMJDIZZY        BMJREFLX        BMJMOOD BMJHALL
U       BMJAPTHY        BMJJOB  BMJDRIVE        BMJILLEG        BMJCRIME
BMJEXPNS        BMJDK   BMJMISC BMJREF1 MJHIGH  MJDRHIGH        MJOTHDR MJPUFFS
MJDRPUFF        MJOTHPUF        MJINVOLV        MJCAREMR        MJCRMORE
MJOTHMOR        MJCARELS        MJCRLESS        MJOTHLES        MJWKEND MJCRWKE
N       MJOTHWKN        ALHIGH  ALDRHIGH        ALOTHDR ALSOME  ALDRSOME
ALOTHSOM        ALOTHDRK        ALYOUDRK        CLOSFRNS        FRNSHER FRNSEX
FRNAGE  FRNTRYH FRNRECH SEENUSE CONFESS TESTMNY TRACKMRK        ARREST  UNPRREF
UNPRREP UNPRBEH UNPROTH AMBULANC        DETECOTH        GIVESELL        TREATMN
T       OTHKNOW LVDHEREA        LVDHEREB        EVRLIVEA        AGEINA1 AGEOUTA
1       AGEINA2 AGEOUTA2        AGEINA3 AGEOUTA3        ALLLIFEA        EVRLIVE
B       AGEINB1 AGEOUTB1        AGEINB2 AGEOUTB2        AGEINB3 AGEOUTB3
ALLLIFEB        EVRLIVEC        AGEINC1 AGEOUTC1        AGEINC2 AGEOUTC2
AGEINC3 AGEOUTC3        ALLLIFEC        SEX     RESPAGE HISPANIC        HISPGRP
RESPRACE        RAGEGRP ENRLCOLL        TYPESCHL        STUDFTPT        EDUC
TOTPEOP UNDAGE18        UNDAGE6 AGE612  AGE1217 HHPAREN NUMPAREN        HHSPOUS
NUMSPOUS        HHSIBLN NUMSIBLN        HHOTREL NUMOTREL        HHFRNDS NUMFRND
S       HHOTPER NUMOTPER        MARITAL EMPLOYED        ROCCUP2 NOLABOR CWE
CWEOCC2 INCOME  ESTHHIN YTHSTUD YSTDFTPT        YTHEDUC YTOTPEOP        MOTHER
FATHER  OLDSIBS NUMOSIBS        YNGSIBS NUMYSIBS        YTHOTREL        NUMYORE
L       YTHOTPER        NUMYOPER        OTHSIBS YTHEMPLD        YTHOCCU2
YNOLABOR        HHAREA  MILINSTA        LOGCAMP COLLEGE RESORT  CONSTR  RANCH
MIGRANTS        TEMPRES HHTYPE  UNDINT  COOPINT PRIVACY ADULTYTH        PAREXAM
Q       ADLTQCD QUEXTYPE        INTVLEN FIID    TOTHHVIS        FINLRES1
VSADLTCM        PHADLTCM        FINLRES2        VSYTHCM PHYTHCM YTHINHH RES1825
```

```
RES2649 RES50OVR        AGR1REL1        AGR1SEX1        AGR1AGE1        AGR1RSP
1       AGR1REL2        AGR1SEX2        AGR1AGE2        AGR1RSP2        AGR1REL
3       AGR1SEX3        AGR1AGE3        AGR1RSP3        AGR1REL4        AGR1SEX
4       AGR1AGE4        AGR1RSP4        AGR2REL1        AGR2SEX1        AGR2AGE
1       AGR2RSP1        AGR2REL2        AGR2SEX2        AGR2AGE2        AGR2RSP
2       AGR2REL3        AGR2SEX3        AGR2AGE3        AGR2RSP3        AGR2REL
4       AGR2SEX4        AGR2AGE4        AGR2RSP4        AGR3REL1        AGR3SEX
1       AGR3AGE1        AGR3RSP1        AGR3REL2        AGR3SEX2        AGR3AGE
2       AGR3RSP2        AGR3REL3        AGR3SEX3        AGR3AGE3        AGR3RSP
3       AGR3REL4        AGR3SEX4        AGR3AGE4        AGR3RSP4        YTH1217
YTH1REL YTH1SEX YTH1AGE YTH1RSP YTH2REL YTH2SEX YTH2AGE YTH2RSP YTH3REL YTH3SEX
YTH3AGE YTH3RSP YTH4REL YTH4SEX YTH4AGE YTH4RSP REGION  DIVISION        POPDENX
IRAGE   IIAGE   IRSEX   IISEX   IRRACEX IIRACEX IRHOIND IIHOIND IRHOGRP IIHOGRP
IRMARIT IIMARIT IREDUC  IIEDUC  IRALCRC IIALCRC IRMJRC  IIMJRC  IRCOCRC IICOCRC
IRSEDRC IISEDRC IRTRANRC        IITRANRC        IRSTIMRC        IISTIMRC
IRANALRC        IIANALRC        IRCIGRC IICIGRC IRINHRC IIINHRC IRHALLRC
IIHALLRC        IRHERRC IIHERRC CATAGE  CATAG2  CATAG3  RACE    HISPRACE
EDUCCAT2        HALFLAG HALYR   HALMON  STMFLAG STMYR   STMMON  SEDFLAG SEDYR
SEDMON  TRQFLAG TRQYR   TRQMON  ANLFLAG ANLYR   ANLMON  ALCFLAG ALCYR   ALCMON
CIGFLAG CIGYR   CIGMON  HERFLAG HERYR   HERMON  MRJFLAG MRJYR   MRJMON  COCFLAG
COCYR   COCMON  INHFLAG INHYR   INHMON  PSYFLAG2        PSYYR2  PSYMON2 SUMFLAG
SUMYR   SUMMON  MJOFLAG MJOYR2  MJOMON2 IEMFLAG IEMYR   IEMMON  VESTR   VEREP
ANALWT  CANALWT NANALWT INITWT  WT1     WT2     CINITWT CWT1    CWT2    NINITWT
NWT1    NWT2
1       1214    63      151     2040    3       16      1       4       99
1       1       1       1       1       1       1       1       1       1
1       1       1       1       1       1       0       2       2       2
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      2       2       2       91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      2
2       2       91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      2       2       2       91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      7       25      1
3       1       3       1       1       1       1       2       91      91
91      91      91      91      91      1       1       1       1       1
98      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      2       91      1       1       1
1       91      91      91      91      91      91      1       91      91
1       1       1       1       91      91      91      91      91      91
1       1       1       1       2       91      91      91      91      91
91      91      91      91      1       1       1       1       91      91
1       91      2       3       91      91      91      0       1       0
0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       1       0       0       0
0       4       99      99      4       99      99      5       4       99
99      4       99      99      4       99      99      1       2       2
1       2       2       1       1       93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      25      93      1       99      99
99      99      99      99      99      1       99      99      99      99
99      99      99      2       98      98      98      98      98      98
1       2       70      2       98      5       3       2       99      99
4       2       2       99      99      99      2       98      1       1
2       98      1       1       2       98      2       98      1       2
```

```
999     1       1       4       98      98      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      993     93      3       2       2       1       1       2
2       2       2       1       1       1       93      1       93      1
1       55      8329    1       1       1       1       98      98      98
2       0       1       2       98      98      98      98      98      98
98      98      98      98      98      98      98      98      98      98
2       1       36      0       98      98      98      98      98      98
98      98      98      98      98      98      4       1       70      0
1       2       70      1       98      98      998     98      98      98
998     98      98      98      98      98      98      98      98      98
98      98      98      98      98      98      98      98      98      1
1       1       70      1       2       1       5       1       2       1
9       9       1       1       4       1       1       1       9       1
9       1       9       1       9       1       9       1       9       1
4       1       9       1       9       1       9       1       4       3
5       1       3       2       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       1
1       1       1       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       7982001
1       82468.5995      101423.3097     0       3.272   24964.4132      1.0096
3.272   31239.2811      .9923   0       24964.1683      1.0105
2       1478    63      151     5931    3       14      2       19      99
1       1       1       1       1       1       1       1       1       1
1       1       1       1       1       1       0       1       1       1
6       99      99      99      99      0       0       0       0       0
0       0       0       0       0       0       0       0       1       0
0       0       0       0       0       0       0       0       0       0
0       0       98      2       2       2       91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      1
1       1       6       99      99      99      99      1       0       0
1       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0
20      2       2       2       91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      7       17      2
5       2       1       1       1       1       1       1       17      7
19      3       2       2       3       1       1       1       1       2
98      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      2       91      1       1       1
1       19      91      91      91      91      91      1       91      91
1       1       1       1       25      91      91      91      91      91
1       1       1       1       1       91      91      91      91      91
91      91      91      91      1       1       95      1       91      91
98      91      1       2       3       4       91      0       0       0
0       1       1       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       1       0
0       1       0       0       0       0       0       0       0       0
0       98      1       1       98      94      94      1       98      1
1       98      2       1       98      2       2       3       99      99
3       99      99      14      14      93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      4       93      2       5       11
98      98      98      98      98      2       0       4       18      28
98      98      98      2       10      17      98      98      98      98
98      2       28      2       98      5       2       2       99      99
7       2       2       99      99      99      2       98      2       98
2       98      2       98      1       2       2       98      5       1
4       99      4       999     9       98      93      93      93      93
```

```
93      93      93      93      93      93      93      93      93      93
93      93      993     93      5       2       2       1       1       2
2       2       2       3       1       1       93      1       93      1
1       55      8329    2       1       2       1       98      98      98
2       1       2       0       5       1       23      0       98      98
98      98      98      98      98      98      98      98      98      98
5       2       28      0       1       2       27      1       98      98
98      98      98      98      98      98      98      98      998     98
98      98      998     98      98      98      998     98      98      98
998     98      98      98      98      98      98      98      98      98
98      98      98      98      98      98      98      98      98      1
1       1       28      1       2       1       5       1       2       1
9       9       5       1       7       1       2       1       2       1
9       1       6       1       9       1       6       1       9       1
4       3       9       1       9       1       9       1       3       3
3       1       3       4       0       0       0       1       0       0
1       0       0       0       0       0       0       0       0       1
1       1       1       0       0       0       0       0       1       1
1       0       0       0       0       0       0       1       0       0
1       1       1       0       1       1       1       0       0       7982001
1       40445.6499      51980.5608      0       1.5764  24964.4132      1.0278
1.5764  31239.2811      1.0555  0       24964.1683      .8663
3       1608    63      151     4771    3       17      2       19      99
1       1       1       1       1       1       1       1       1       1
1       1       1       1       1       1       0       2       2       2
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      2       2       2       91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      2
2       2       91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      2       2       2       91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      94      94      17
2       1       2       1       1       94      1       1       17      91
91      91      91      91      91      1       1       98      1       2
98      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      2       91      1       1       1
1       91      91      91      91      91      91      1       91      91
1       1       1       1       91      91      91      91      91      91
1       1       1       1       2       91      91      91      91      91
91      91      91      91      1       1       1       1       91      91
9       91      2       1       91      91      91      0       0       0
0       0       0       0       0       1       0       0       0       1
0       0       0       0       0       1       0       1       0       0
0       0       0       0       0       0       0       0       0       0
0       98      2       2       98      2       2       1       98      94
2       98      2       2       98      2       2       98      2       1
98      2       2       2       1       93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      15      93      1       99      99
99      99      99      99      99      2       98      98      98      98
98      98      1       99      99      99      99      99      99      99
99      2       21      2       98      5       1       2       99      99
6       3       2       99      99      99      1       2       2       98
1       1       2       98      2       98      2       98      5       1
5       99      1       1       98      98      93      93      93      93
93      93      93      93      93      93      93      93      93      93
```

```
93      93      993     93      5       2       2       1       2       2
2       2       2       1       1       1       93      1       93      1
1       60      8329    2       1       2       1       98      98      98
2       1       2       0       2       2       21      1       98      98
98      98      98      98      98      98      98      98      98      98
4       1       47      0       1       2       42      0       98      98
98      98      98      98      98      98      98      98      998     98
98      98      998     98      98      98      998     98      98      98
998     98      98      98      98      98      98      98      98      98
98      98      98      98      98      98      98      98      98      1
1       1       21      1       2       1       5       1       2       1
9       9       5       1       6       1       1       3       9       1
9       1       9       1       9       1       9       1       9       1
2       3       9       1       9       1       9       1       2       2
2       1       3       3       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       1
1       1       1       1       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       7982001
1       12963.4172      15711.5542      0       .5309   24964.4132      .978
.5309   31239.2811      .9473   0       24964.1683      1.0459
4       1661    63      151     292     2       91      91      91      91
1       1       1       1       1       1       1       1       1       1
1       1       1       1       1       1       0       2       2       2
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      2       2       2       91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      2
2       2       91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      2       2       2       91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      7       7       1
1       1       1       1       1       1       1       2       91      91
91      91      91      91      91      1       1       1       1       2
98      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      2       91      1       1       1
1       91      91      91      91      91      91      1       91      91
1       1       1       1       91      91      91      91      91      91
1       1       1       1       2       91      91      91      91      91
91      91      91      91      1       1       1       1       91      91
91      91      1       2       91      91      91      0       0       0
0       0       1       0       0       0       0       0       0       0
0       0       0       0       0       1       0       1       0       0
0       0       0       0       0       0       0       0       0       0
0       98      98      98      98      98      98      3       3       99
99      3       99      99      2       99      99      4       99      99
3       99      99      1       1       93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      4       93      1       99      99
99      99      99      99      99      2       43      45      98      98
98      98      98      4       0       43      98      98      98      98
98      2       45      2       98      5       2       2       99      99
4       4       1       0       1       2       2       98      1       1
2       98      2       98      2       98      2       98      1       2
999     98      1       4       8       98      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      993     93      5       2       2       1       1       2
```

```
2       2       2       1       1       1       93      1       93      1
1       60      8329    2       1       2       1       1       2       1
1       0       2       0       98      98      98      98      98      98
98      98      98      98      98      98      98      98      98      98
4       1       45      0       1       2       45      1       98      98
98      98      98      98      98      98      98      98      998     98
98      98      998     98      98      98      998     98      98      98
998     98      2       2       2       17      0       2       2       15
1       98      98      98      98      98      98      98      98      1
1       1       45      1       2       1       5       1       2       1
9       9       1       1       4       1       1       1       9       1
9       1       9       1       9       1       9       1       9       1
9       1       9       1       9       1       9       1       4       3
4       1       3       2       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       1
1       1       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       7982001
1       54979.0503      67615.5202      0       2.1813  24964.4132      1.0096
2.1813  31239.2811      .9923   0       24964.1683      1.0105
5       1803    63      151     5232    2       15      2       19      99
1       1       1       1       1       1       1       0       1       0
0       0       0       0       1       1       0       2       2       2
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      2       2       2       91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      2
2       2       91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      2       2       2       91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      7       18      1
4       1       1       1       1       1       1       1       30      91
91      91      91      91      91      1       1       1       1       2
98      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      2       91      1       1       1
1       91      91      91      91      91      91      1       91      91
1       1       1       1       91      91      91      91      91      91
1       1       1       1       2       91      91      91      91      91
91      91      91      91      1       1       1       1       91      91
1       91      2       3       91      91      91      0       0       0
0       1       0       1       0       0       0       0       0       0
0       0       0       0       0       0       0       0       1       0
0       1       0       0       0       0       0       0       0       0
0       3       99      99      3       99      99      1       98      94
2       98      94      2       98      2       2       98      2       1
98      2       2       3       2       93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      11      93      1       99      99
99      99      99      99      99      2       98      98      98      98
98      98      1       99      99      99      99      99      99      99
99      2       36      2       98      5       2       2       99      99
6       3       1       0       2       0       2       98      1       1
2       98      2       98      2       98      2       98      1       2
999     1       1       2       98      98      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      993     93      5       2       2       1       1       2
2       2       2       1       1       1       93      1       93      1
```

```
1       60      8329    3       1       3       1       98      98      98
2       0       2       0       98      98      98      98      98      98
98      98      98      98      98      98      98      98      98      98
4       1       37      0       1       2       36      1       98      98
98      98      98      98      98      98      98      98      998     98
98      98      998     98      98      98      998     98      98      98
998     98      98      98      98      98      98      98      98      98
98      98      98      98      98      98      98      98      98      1
1       1       36      1       2       1       5       1       2       1
9       9       1       1       6       1       1       1       9       1
9       1       9       1       9       1       9       1       9       1
4       3       9       1       9       1       9       1       4       3
4       1       3       3       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       1
1       1       1       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       7982001
1       54979.0503      67615.5202      0       2.1813  24964.4132      1.0096
2.1813  31239.2811      .9923   0       24964.1683      1.0105
6       2173    63      151     5752    2       12      1       4       99
1       1       1       1       1       1       1       1       1       1
1       1       1       1       1       1       0       2       2       2
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      1       2       1       4       99      99      99
99      1       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       35      2
2       2       91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      2       2       2       91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      7       14      1
3       1       3       1       1       1       1       1       30      7
30      3       6       0       3       1       1       98      1       1
98      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      2       91      1       1       1
1       91      91      91      91      91      91      1       91      91
1       1       1       1       91      91      91      91      91      91
1       1       1       1       1       91      91      91      91      91
91      91      91      91      1       1       1       1       91      91
1       91      2       3       4       5       91      0       0       0
0       0       1       1       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0
0       0       0       0       1       0       0       0       0       0
0       1       1       1       1       1       1       3       1       1
1       1       2       2       1       2       2       1       1       1
1       2       2       2       2       93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      18      93      1       99      99
99      99      99      99      99      2       18      36      98      98
98      98      98      2       0       18      98      98      98      98
98      2       36      2       98      5       2       2       99      99
98      4       1       0       0       1       1       1       1       1
2       98      1       1       2       98      2       98      1       2
999     98      1       1       7       98      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      993     93      5       2       2       1       1       2
2       2       2       1       1       1       93      1       93      1
1       55      8329    2       1       2       1       1       2       1
```

```
1       1       2       1       2       1       18      0       98      98
98      98      98      98      98      98      98      98      98      98
4       1       39      0       1       2       36      1       98      98
98      98      98      98      98      98      7       1       64      0
98      98      998     98      98      98      998     98      98      98
998     98      1       2       2       14      1       98      98      98
98      98      98      98      98      98      98      98      98      1
1       1       36      1       2       1       5       1       2       1
9       9       1       1       4       2       1       1       6       1
9       1       9       1       4       1       9       1       9       1
4       1       9       1       9       1       9       1       4       3
4       1       3       2       0       0       0       0       0       0
0       0       0       1       1       0       0       0       0       1
1       1       1       0       0       0       0       0       1       0
0       0       0       0       0       0       0       1       1       0
1       1       0       0       0       0       1       1       0       7982001
1       54979.0503      67615.5202      0       2.1813  24964.4132      1.0096
2.1813  31239.2811      .9923   0       24964.1683      1.0105
7       6019    63      151     4519    2       16      1       1       5
1       1       1       1       1       1       1       1       1       1
1       1       1       1       1       1       0       2       2       2
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      2       2       2       91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      2
2       2       91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      2       2       2       91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      7       19      1
4       2       1       1       1       1       1       1       20      91
91      91      91      91      91      1       1       1       1       2
98      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      94      91      1       1       1
1       91      91      91      91      91      91      1       91      91
1       1       1       1       91      91      91      91      91      91
1       1       1       1       1       91      91      91      91      91
91      91      91      91      1       1       1       1       91      91
1       91      2       3       91      91      91      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      4       1       2       2       3
3       0       1       0       0       0       0       0       0       0
0       0       0       0       2       3       93      1       99      99
99      99      99      99      99      2       18      22      98      98
98      98      98      2       28      32      98      98      98      98
98      2       32      2       98      5       2       2       99      99
4       3       1       2       0       0       2       98      1       1
2       98      2       98      2       98      2       98      1       2
999     1       1       11      5       98      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      993     93      3       2       2       1       1       2
2       2       2       1       1       1       93      1       93      2
2       60      8329    2       1       2       1       98      98      98
2       0       2       0       98      98      98      98      98      98
```

```
98      98      98      98      98      98      98      98      98      98
4       1       32      0       1       2       32      1       98      98
98      98      98      98      98      98      98      98      998     98
98      98      998     98      98      98      998     98      98      98
998     98      98      98      98      98      98      98      98      98
98      98      98      98      98      98      98      98      98      1
1       1       32      1       2       1       5       1       2       1
9       9       1       1       4       1       1       1       9       1
9       1       9       1       9       1       9       1       9       1
1       1       9       1       9       1       9       1       3       3
3       1       3       2       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       1
1       1       1       1       1       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       7982001
1       40445.6499      0       182263.0098     1.5764  24964.4132      1.0278
0       31239.2811      1.0555  8.4274  24964.1683      .8663
8       6098    63      151     3453    2       91      91      91      91
1       1       1       1       1       1       1       1       1       0
1       1       1       1       0       1       0       2       2       2
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      2       2       2       91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      2
2       2       91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      2       2       2       91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      94      4       17
1       1       1       1       1       1       1       1       16      94
17      2       4       0       1       1       1       1       1       2
98      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      2       91      1       1       1
1       17      91      91      91      91      91      1       91      91
1       1       1       1       18      91      91      91      91      91
1       1       1       1       1       91      91      91      91      91
91      91      91      91      1       1       1       1       91      91
91      91      1       3       2       91      91      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      4       0       99      99      99
99      99      99      99      99      99      99      99      99      99
99      99      99      99      99      6       93      2       1       13
98      98      98      98      98      2       13      20      98      98
98      98      98      1       99      99      99      99      99      99
99      2       20      2       98      5       1       1       1       1
6       4       2       99      99      99      1       2       2       98
1       2       2       98      2       98      2       98      5       1
15      99      1       1       7       98      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      993     93      5       2       2       1       1       2
2       2       2       1       1       1       93      1       93      2
2       45      8329    3       1       2       1       1       3       1
1       1       1       1       2       2       20      1       98      98
98      98      98      98      98      98      98      98      98      98
```

```
4       2       48      0       98      98      98      98      98      98
98      98      98      98      98      98      1       1       50      0
98      98      998     98      98      98      998     98      98      98
998     98      2       2       1       16      1       2       1       13
0       98      98      98      98      98      98      98      98      1
1       1       20      1       2       1       5       1       2       1
9       9       5       1       6       1       1       3       4       1
9       1       9       1       9       1       9       1       9       1
9       1       9       1       9       1       9       1       2       2
2       1       3       3       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       1
1       1       0       0       0       0       0       0       1       1
0       0       0       0       0       0       0       0       0       0
1       1       0       1       1       0       0       0       0       7982001
1       12963.4172      0       74114.0009      .5309   24964.4132      .978
0       31239.2811      .9473   2.8385  24964.1683      1.0459
9       6149    63      151     3251    2       18      2       19      99
1       1       1       1       1       1       1       1       1       1
1       1       1       1       1       1       0       2       2       2
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      2       2       2       91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      2
2       2       91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      2       2       2       91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      7       17      1
2       1       2       1       1       1       1       2       91      91
91      91      91      91      91      1       1       1       1       98
98      91      91      91      91      91      91      91      91      91
91      91      91      91      91      91      91      91      91      91
91      91      91      91      91      1       91      1       1       1
1       91      91      91      91      91      91      2       91      91
1       1       95      1       91      91      91      91      91      91
1       1       95      1       2       91      91      91      91      91
91      91      91      91      1       1       1       1       91      91
1       91      2       3       91      91      91      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      93      93      93      6       0       99      99      99
99      99      99      99      99      99      99      99      99      99
99      99      99      99      99      98      2       1       99      99
99      99      99      99      99      2       98      98      98      98
98      98      1       99      99      99      99      99      99      99
99      2       48      2       98      5       2       2       99      99
7       1       2       99      99      99      2       98      1       1
2       98      2       98      2       98      2       98      1       1
1       99      1       1       98      98      93      93      93      93
93      93      93      93      93      93      93      93      93      93
93      93      993     93      5       2       2       1       1       2
2       2       2       1       1       1       93      1       93      2
2       45      8329    2       1       2       1       98      98      98
2       0       2       0       98      98      98      98      98      98
98      98      98      98      98      98      98      98      98      98
4       1       49      0       1       2       48      1       98      98
```

```
98      98      98      98      98      98      98      98      998     98
98      98      998     98      98      98      998     98      98      98
998     98      98      98      98      98      98      98      98      98
98      98      98      98      98      98      98      98      98      1
1       1       48      1       2       1       5       1       2       1
9       9       1       1       7       1       1       1       9       1
9       1       9       1       9       1       9       1       9       1
2       3       9       1       9       1       9       1       4       3
4       1       3       4       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       1
1       1       1       1       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       0
0       0       0       0       0       0       0       0       0       7982001
1       54979.0503      0       294183.1968     2.1813  24964.4132      1.0096
0       31239.2811      .9923   11.6614 24964.1683      1.0105
download failed: s3://ads-508-azhang/finalproject/NHSDA-1979-DS0001-data-excel.
tsv to - [Errno 32] Broken pipe
```

In [54]:
```python
s3_client = boto3.client("s3")
```

In [104]:
```python
bucket = 'ads-508-azhang'
key = 'finalproject/NHSDA-1988-DS0001-data-excel.tsv'
```

In [105]:
```python
response = s3_client.get_object(Bucket = bucket, Key = key)
```

In [106]:
```python
df = pd.read_csv(response.get("Body"))
```

In [107]:
```python
df.head(1)
```

Out[107]:

	CASEID\tRESPID\tENCPSU\tENCSEG\tENCCASE\tCIGTRY\tCIG5PK\tCIGREC\tAVCIG\tCIGTIME\tCIGIN
0	

In [63]:
```python
#Create Athena DB Schema
```

In [64]:
```python
import boto3
import sagemaker

sess = sagemaker.Session()
bucket = sess.default_bucket()
role = sagemaker.get_execution_role()
region = boto3.Session().region_name
```

In [65]:
```python
ingest_create_athena_db_passed = False
```

In [66]:
```python
get_ipython().run_line_magic('store', '-r s3_public_path_tsv')
```

```
In [67]: try:
             s3_public_path_tsv
         except NameError:
             print("********************************************************************
         *********")
             print("[ERROR] PLEASE RE-RUN THE PREVIOUS COPY TSV TO S3 NOTEBOOK *********
         *********")
             print("[ERROR] THIS NOTEBOOK WILL NOT RUN PROPERLY. ***********************
         *********")
             print("********************************************************************
         *********")
```

```
In [68]: print(s3_public_path_tsv)
```

```
s3://ads-508-azhang/finalproject/
```

```
In [69]: get_ipython().run_line_magic('store', '-r s3_private_path_tsv')
```

```
In [70]: try:
             s3_private_path_tsv
         except NameError:
             print("********************************************************************
         *********")
             print("[ERROR] PLEASE RE-RUN THE PREVIOUS COPY TSV TO S3 NOTEBOOK *********
         *********")
             print("[ERROR] THIS NOTEBOOK WILL NOT RUN PROPERLY. ***********************
         *********")
             print("********************************************************************
         *********")
```

```
In [71]: print(s3_private_path_tsv)
```

```
s3://sagemaker-us-east-1-189468192453/ads-508-azhang/finalproject
```

```
In [72]: #import PyAthena
         get_ipython().system('pip install --disable-pip-version-check -q PyAthena==2.1.
         0')
         from pyathena import connect
```

```
/opt/conda/lib/python3.7/site-packages/secretstorage/dhcrypto.py:16: Cryptograp
hyDeprecationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
/opt/conda/lib/python3.7/site-packages/secretstorage/util.py:25: CryptographyDe
precationWarning: int_from_bytes is deprecated, use int.from_bytes instead
  from cryptography.utils import int_from_bytes
WARNING: Running pip as the 'root' user can result in broken permissions and co
nflicting behaviour with the system package manager. It is recommended to use a
virtual environment instead: https://pip.pypa.io/warnings/venv
```

```
In [73]: database_name = "drugs"
```

```
In [110]: # Set S3 staging directory -- this is a temporary directory used for Athena que
          ries
          s3_staging_dir = "s3://{0}/ads-508-azhang/finalproject/staging".format(bucket)
```

```
In [111]: conn = connect(region_name=region, s3_staging_dir=s3_staging_dir)
```

In [112]:

```
statement = "CREATE DATABASE IF NOT EXISTS {}".format(database_name)
print(statement)
```

```
CREATE DATABASE IF NOT EXISTS drugs
```

In [113]:

```
import pandas as pd

pd.read_sql(statement, conn)
```

Out[113]:

```
—
```

In [114]:

```
statement = "SHOW DATABASES"

df_show = pd.read_sql(statement, conn)
df_show.head(5)
```

Out[114]:

	database_name
0	default
1	drugs
2	dsoaws

In [119]:

```
drug_dir = 's3://sagemaker-us-east-1-189468192453/ads-508-azhang/finalproject'
```

In [120]:

```
table_name ='NHSDA_1979'
pd.read_sql(f'DROP TABLE IF EXISTS {database_name}.{table_name}', conn)
file_name1 = 'NHSDA-1979-DS0001-data-excel.tsv'
file_name2 = 'NHSDA-1988-DS0001-data-excel.tsv'
file_name3 = 'NHSDA-1995-DS0001-data-excel.tsv'

create_table = f"""
CREATE EXTERNAL TABLE IF NOT EXISTS {database_name}.{table_name}(
                CASEID  float,
RESPID  float,
ENCPSU  float,
ENCSEG  float,
ENCCASE  float,
CIGMORLS  float,
CIGTRY  float,
CIG5PK  float,
CIGREC  float,
AVCIG  float,
HRDHER  float,
HRDMJ  float,
HRDCOC  float,
HRDLSD  float,
HRDBAR  float,
HRDTRN  float,
HRDAMP  float,
ADDHER  float,
ADDALC  float,
ADDMJ  float,
ADDTOB  float,
ADDBAR  float,
ADDTRN  float,
ADDAMP  float,
ADDLSD  float,
ADDCOC  float,
ADDNONE  float,
SEDLIKE  float,
SEDFEEL  float,
SEDNEED  float,
SEDREC  float,
SED30MOA  float,
SED30MOB  float,
SED30MOC  float,
SEDDAL30  float,
BUTISOL  float,
BUTICAPS  float,
AMYTAL  float,
ESKABARB  float,
LUMINAL  float,
MEBARAL  float,
AMOBARB  float,
PHENOBAR  float,
ALURATE  float,
PLACIDYL  float,
DORIDEN  float,
NOLUDAR  float,
SOPOR  float,
QUAALUDE  float,
PAREST  float,
NOCTEC  float,
METHAQ  float,
CHHYD  float,
NEMBUTAL  float,
CARBTAL  float,
SECONAL  float,
TUINAL  float,
PENTOB  float,
SECOB  float,
DALMANE  float,
SEDDKNAM  float,
NOSEDAT  float,
SEDAGE  float,
TRNLIKE  float,
TRNFEEL  float,
TRNNEED  float,
TRANREC  float,
TRN30MOA  float,
TRN30MOB  float,
TRN30MOC  float,
TRNBEN30  float,
VALIUM  float,
LIBRIUM  float,
LIBRITAB  float,
SKLY  float,
SERAX  float,
TRANXENE  float,
ATIVAN  float,
VERSTRAN  float,
MEPRSPAN  float,
MILTOWN  float,
EQUANIL  float,
MEPROB  float,
VISTAR  float,
ATARAX  float,
BENADRYL  float,
TRDKNAM  float,
NOTRANQ  float,
TRANAGE  float,
STIMLIKE  float,
STIMFEEL  float,
STIMNEED  float,
STIMREC  float,
STM30MOA  float,
STM30MOB  float,
STMRIT30  float,
STMCYL30  float,
DEXED  float,
DEXAMYL  float,
ESKAT  float,
BENZ  float,
BIPHET  float,
DESOXYN  float,
DETAMP  float,
METHI  float,
OBLA  float,
TENUATE  float,
TEPANIL  float,
DIDREX  float,
PLEGINE  float,
PRELUDIN  float,
PRESATE  float,
IONAMIN  float,
PONDIMIN  float,
VORANIL  float,
SANOREX  float,
RITALIN  float,
CYLERT  float,
STMDKNAM  float,
NOSTIMS  float,
STIMAGE  float,
ANALLIKE  float,
ANALFEEL  float,
ANALNEED  float,
ANALREC  float,
ANL30MOA  float,
ANL30MOB  float,
ANL30MOC  float,
ANLTAL30  float,
DARVON  float,
DOLENE  float,
SK65A  float,
PROPOXY  float,
LERITINE  float,
LEVODRO  float,
PERCODAN  float,
DEMEROL  float,
DILAUD  float,
TYLCOD  float,
CODEINE  float,
DOLOP  float,
WESTODON  float,
METHDON  float,
TALWIN  float,
ANLDKNAM  float,
ANALNONE  float,
ANALAGE  float,
ALCFIRST  float,
ALCTRY  float,
ALCREC  float,
ALCDAYS  float,
MODR30A  float,
MODR30DY  float,
UNDSTAS1  float,
VRA7AS1  float,
MRKEAAS1  float,
VRA8AS1  float,
MJKNOWN  float,
MJOPP  float,
MJFIRST  float,
MJAGE  float,
MJLIVE  float,
MJREC  float,
MJDAY30A  float,
MJTOT  float,
UNDSTAS2  float,
VRM9AS2  float,
MRKEAAS2  float,
VRM10AS2  float,
INHREAD  float,
INHOPP  float,
INHFIRST  float,
INHAGE  float,
GAS  float,
SPPAINT  float,
AEROS  float,
GLUE  float,
SOLVENT  float,
AMYLNIT  float,
ETHER  float,
NITOXID  float,
ODORIZER  float,
INHNEVER  float,
GAS30A  float,
SPPAN30A  float,
AEROS30A  float,
GLUE30A  float,
SOLVN30A  float,
AMLNT30A  float,
ETHER30A  float,
NOX30A  float,
ODR30A  float,
INH30NO  float,
INHREC  float,
INHTOT  float,
INHODRHR  float,
INHODRUS  float,
UNDSTAS3  float,
VRG10AS3  float,
MRKEAAS3  float,
VRG11AS3  float,
HALLOPP  float,
HALFIRST  float,
HALLAGE  float,
HALLREC  float,
HAL30USE  float,
HALLTOT  float,
HALPCPHR  float,
PCP  float,
HALPCP30  float,
UNDSTAS4  float,
VRL10AS4  float,
MRKEAAS4  float,
VRL11AS4  float,
COCOPP  float,
COCFIRST  float,
COCAGE  float,
COCREC  float,
COCUS30A  float,
COCTOT  float,
UNDSTAS5  float,
VRC7AS5  float,
MRKEAAS5  float,
VRC8AS5  float,
HERKNOW  float,
HEROPP  float,
HERFIRST  float,
HERAGE  float,
HERREC  float,
HER30USE  float,
HERTOT  float,
HERFRNDS  float,
HERNOADR  float,
HERNEEDL  float,
UNDSTAS6  float,
VRH11AS6  float,
MRKEAAS6  float,
VRH12AS6  float,
SPLCOC  float,
SPLHAL  float,
SPLCIG  float,
SPLHER  float,
SPLBEER  float,
SPLLQR  float,
SPLMJR  float,
SPLPILLS  float,
SPLINH  float,
GMJNOHO  float,
GMJNONE  float,
GMJMED  float,
GMJJOB  float,
GMJFUN  float,
GMJRELAX  float,
GMJAWARE  float,
GMJCNFDN  float,
GMJDEAL  float,
GMJSLEEP  float,
GMJSEX  float,
GMJAPPET  float,
GMJDK  float,
GMJMISC  float,
GMJREF1  float,
BMJCONTR  float,
BMJMEMRY  float,
BMJNONE  float,
BMJHABIT  float,
BMJSTRGR  float,
BMJHLTH  float,
BMJDIZZY  float,
BMJREFLX  float,
BMJMOOD  float,
BMJHALLU  float,
BMJAPTHY  float,
BMJJOB  float,
BMJDRIVE  float,
BMJILLEG  float,
BMJCRIME  float,
BMJEXPNS  float,
BMJDK  float,
BMJMISC  float,
BMJREF1  float,
MJHIGH  float,
MJDRHIGH  float,
MJOTHDR  float,
MJPUFFS  float,
MJDRPUFF  float,
MJOTHPUF  float,
MJINVOLV  float,
MJCAREMR  float,
MJCRMORE  float,
MJOTHMOR  float,
MJCARELS  float,
MJCRLESS  float,
MJOTHLES  float,
MJWKEND  float,
MJCRWKEN  float,
MJOTHWKN  float,
ALHIGH  float,
ALDRHIGH  float,
ALOTHDR  float,
ALSOME  float,
ALDRSOME  float,
ALOTHSOM  float,
ALOTHDRK  float,
ALYOUDRK  float,
CLOSFRNS  float,
FRNSHER  float,
FRNSEX  float,
FRNAGE  float,
FRNTRYH  float,
FRNRECH  float,
SEENUSE  float,
CONFESS  float,
TESTMNY  float,
TRACKMRK  float,
ARREST  float,
UNPRREF  float,
UNPRREP  float,
UNPRBEH  float,
UNPROTH  float,
AMBULANC  float,
DETECOTH  float,
GIVESELL  float,
TREATMNT  float,
OTHKNOW  float,
LVDHEREA  float,
LVDHEREB  float,
EVRLIVEA  float,
AGEINA1  float,
AGEOUTA1  float,
AGEINA2  float,
AGEOUTA2  float,
AGEINA3  float,
AGEOUTA3  float,
ALLLIFEA  float,
EVRLIVEB  float,
AGEINB1  float,
AGEOUTB1  float,
AGEINB2  float,
AGEOUTB2  float,
AGEINB3  float,
AGEOUTB3  float,
ALLLIFEB  float,
EVRLIVEC  float,
AGEINC1  float,
AGEOUTC1  float,
AGEINC2  float,
AGEOUTC2  float,
AGEINC3  float,
AGEOUTC3  float,
ALLLIFEC  float,
SEX  float,
RESPAGE  float,
HISPANIC  float,
HISPGRP  float,
RESPRACE  float,
RAGEGRP  float,
ENRLCOLL  float,
TYPESCHL  float,
STUDFTPT  float,
EDUC  float,
TOTPEOP  float,
UNDAGE18  float,
UNDAGE6  float,
AGE612  float,
AGE1217  float,
HHPAREN  float,
NUMPAREN  float,
HHSPOUS  float,
NUMSPOUS  float,
HHSIBLN  float,
NUMSIBLN  float,
HHOTREL  float,
NUMOTREL  float,
HHFRNDS  float,
NUMFRNDS  float,
HHOTPER  float,
NUMOTPER  float,
MARITAL  float,
EMPLOYED  float,
ROCCUP2  float,
NOLABOR  float,
CWE  float,
CWEOCC2  float,
INCOME  float,
ESTHHIN  float,
YTHSTUD  float,
YSTDFTPT  float,
YTHEDUC  float,
YTOTPEOP  float,
MOTHER  float,
FATHER  float,
OLDSIBS  float,
NUMOSIBS  float,
YNGSIBS  float,
NUMYSIBS  float,
YTHOTREL  float,
NUMYOREL  float,
YTHOTPER  float,
NUMYOPER  float,
OTHSIBS  float,
YTHEMPLD  float,
YTHOCCU2  float,
YNOLABOR  float,
HHAREA  float,
MILINSTA  float,
LOGCAMP  float,
COLLEGE  float,
RESORT  float,
CONSTR  float,
RANCH  float,
MIGRANTS  float,
TEMPRES  float,
HHTYPE  float,
UNDINT  float,
COOPINT  float,
PRIVACY  float,
ADULTYTH  float,
PAREXAMQ  float,
ADLTQCD  float,
QUEXTYPE  float,
INTVLEN  float,
FIID  float,
TOTHHVIS  float,
FINLRES1  float,
VSADLTCM  float,
PHADLTCM  float,
FINLRES2  float,
VSYTHCM  float,
PHYTHCM  float,
YTHINHH  float,
RES1825  float,
RES2649  float,
RES50OVR  float,
AGR1REL1  float,
AGR1SEX1  float,
AGR1AGE1  float,
AGR1RSP1  float,
AGR1REL2  float,
AGR1SEX2  float,
AGR1AGE2  float,
AGR1RSP2  float,
AGR1REL3  float,
AGR1SEX3  float,
AGR1AGE3  float,
AGR1RSP3  float,
AGR1REL4  float,
AGR1SEX4  float,
AGR1AGE4  float,
AGR1RSP4  float,
AGR2REL1  float,
AGR2SEX1  float,
AGR2AGE1  float,
AGR2RSP1  float,
AGR2REL2  float,
AGR2SEX2  float,
AGR2AGE2  float,
AGR2RSP2  float,
AGR2REL3  float,
AGR2SEX3  float,
AGR2AGE3  float,
AGR2RSP3  float,
AGR2REL4  float,
AGR2SEX4  float,
AGR2AGE4  float,
AGR2RSP4  float,
AGR3REL1  float,
AGR3SEX1  float,
AGR3AGE1  float,
AGR3RSP1  float,
AGR3REL2  float,
AGR3SEX2  float,
AGR3AGE2  float,
AGR3RSP2  float,
AGR3REL3  float,
AGR3SEX3  float,
AGR3AGE3  float,
AGR3RSP3  float,
AGR3REL4  float,
AGR3SEX4  float,
AGR3AGE4  float,
AGR3RSP4  float,
YTH1217  float,
YTH1REL  float,
YTH1SEX  float,
YTH1AGE  float,
YTH1RSP  float,
YTH2REL  float,
YTH2SEX  float,
YTH2AGE  float,
YTH2RSP  float,
YTH3REL  float,
YTH3SEX  float,
YTH3AGE  float,
YTH3RSP  float,
YTH4REL  float,
YTH4SEX  float,
YTH4AGE  float,
YTH4RSP  float,
REGION  float,
DIVISION  float,
POPDENX  float,
IRAGE  float,
IIAGE  float,
IRSEX  float,
IISEX  float,
IRRACEX  float,
IIRACEX  float,
IRHOIND  float,
IIHOIND  float,
IRHOGRP  float,
IIHOGRP  float,
IRMARIT  float,
IIMARIT  float,
IREDUC  float,
IIEDUC  float,
IRALCRC  float,
IIALCRC  float,
IRMJRC  float,
IIMJRC  float,
IRCOCRC  float,
IICOCRC  float,
IRSEDRC  float,
IISEDRC  float,
IRTRANRC  float,
IITRANRC  float,
IRSTIMRC  float,
IISTIMRC  float,
IRANALRC  float,
IIANALRC  float,
IRCIGRC  float,
IICIGRC  float,
IRINHRC  float,
IIINHRC  float,
IRHALLRC  float,
IIHALLRC  float,
IRHERRC  float,
IIHERRC  float,
CATAGE  float,
CATAG2  float,
CATAG3  float,
RACE  float,
HISPRACE  float,
EDUCCAT2  float,
HALFLAG  float,
HALYR  float,
HALMON  float,
STMFLAG  float,
STMYR  float,
STMMON  float,
SEDFLAG  float,
SEDYR  float,
SEDMON  float,
TRQFLAG  float,
TRQYR  float,
TRQMON  float,
ANLFLAG  float,
ANLYR  float,
ANLMON  float,
ALCFLAG  float,
ALCYR  float,
ALCMON  float,
CIGFLAG  float,
CIGYR  float,
CIGMON  float,
HERFLAG  float,
HERYR  float,
HERMON  float,
MRJFLAG  float,
MRJYR  float,
MRJMON  float,
COCFLAG  float,
COCYR  float,
COCMON  float,
INHFLAG  float,
INHYR  float,
INHMON  float,
PSYFLAG2  float,
PSYYR2  float,
PSYMON2  float,
SUMFLAG  float,
SUMYR  float,
SUMMON  float,
MJOFLAG  float,
MJOYR2  float,
MJOMON2  float,
IEMFLAG  float,
IEMYR  float,
IEMMON  float,
VESTR  float,
VEREP  float,
ANALWT float,
CANALWT float,
NANALWT float,
INITWT float,
WT1 float,
WT2 float,
CINITWT float,
CWT1 float,
CWT2 float,
NINITWT float,
NWT1 float,
NWT2 float
)

                ROW FORMAT DELIMITED
                FIELDS TERMINATED BY '	'
                LINES TERMINATED BY '\n'
                LOCATION '{drug_dir}/NHSDA-1979-DS0001-data-excel'
                TBLPROPERTIES ('skip.header.line.count'='1')
"""
```

In [121]:
```
pd.read_sql(create_table, conn)
```

Out[121]:

In [122]:
```
pd.read_sql(f'SELECT count(*) FROM {database_name}.{table_name} LIMIT 5', conn)
```

Out[122]:
```
   _col0
0   7224
```

In [123]:

```
table_name2 ='NHSDA_1988'
pd.read_sql(f'DROP TABLE IF EXISTS {database_name}.{table_name2}', conn)

create_table = f"""
CREATE EXTERNAL TABLE IF NOT EXISTS {database_name}.{table_name2}(
                CASEID  float,
RESPID  float,
ENCPSU  float,
ENCSEG  float,
ENCCASE  float,
CIGMORLS  float,
CIGTRY  float,
CIG5PK  float,
CIGREC  float,
AVCIG  float,
HRDHER  float,
HRDMJ  float,
HRDCOC  float,
HRDLSD  float,
HRDBAR  float,
HRDTRN  float,
HRDAMP  float,
ADDHER  float,
ADDALC  float,
ADDMJ  float,
ADDTOB  float,
ADDBAR  float,
ADDTRN  float,
ADDAMP  float,
ADDLSD  float,
ADDCOC  float,
ADDNONE  float,
SEDLIKE  float,
SEDFEEL  float,
SEDNEED  float,
SEDREC  float,
SED30MOA  float,
SED30MOB  float,
SED30MOC  float,
SEDDAL30  float,
BUTISOL  float,
BUTICAPS  float,
AMYTAL  float,
ESKABARB  float,
LUMINAL  float,
MEBARAL  float,
AMOBARB  float,
PHENOBAR  float,
ALURATE  float,
PLACIDYL  float,
DORIDEN  float,
NOLUDAR  float,
SOPOR  float,
QUAALUDE  float,
PAREST  float,
NOCTEC  float,
METHAQ  float,
CHHYD  float,
NEMBUTAL  float,
CARBTAL  float,
SECONAL  float,
TUINAL  float,
PENTOB  float,
SECOB  float,
DALMANE  float,
SEDDKNAM  float,
NOSEDAT  float,
SEDAGE  float,
TRNLIKE  float,
TRNFEEL  float,
TRNNEED  float,
TRANREC  float,
TRN30MOA  float,
TRN30MOB  float,
TRN30MOC  float,
TRNBEN30  float,
VALIUM  float,
LIBRIUM  float,
LIBRITAB  float,
SKLY  float,
SERAX  float,
TRANXENE  float,
ATIVAN  float,
VERSTRAN  float,
MEPRSPAN  float,
MILTOWN  float,
EQUANIL  float,
MEPROB  float,
VISTAR  float,
ATARAX  float,
BENADRYL  float,
TRDKNAM  float,
NOTRANQ  float,
TRANAGE  float,
STIMLIKE  float,
STIMFEEL  float,
STIMNEED  float,
STIMREC  float,
STM30MOA  float,
STM30MOB  float,
STMRIT30  float,
STMCYL30  float,
DEXED  float,
DEXAMYL  float,
ESKAT  float,
BENZ  float,
BIPHET  float,
DESOXYN  float,
DETAMP  float,
METHI  float,
OBLA  float,
TENUATE  float,
TEPANIL  float,
DIDREX  float,
PLEGINE  float,
PRELUDIN  float,
PRESATE  float,
IONAMIN  float,
PONDIMIN  float,
VORANIL  float,
SANOREX  float,
RITALIN  float,
CYLERT  float,
STMDKNAM  float,
NOSTIMS  float,
STIMAGE  float,
ANALLIKE  float,
ANALFEEL  float,
ANALNEED  float,
ANALREC  float,
ANL30MOA  float,
ANL30MOB  float,
ANL30MOC  float,
ANLTAL30  float,
DARVON  float,
DOLENE  float,
SK65A  float,
PROPOXY  float,
LERITINE  float,
LEVODRO  float,
PERCODAN  float,
DEMEROL  float,
DILAUD  float,
TYLCOD  float,
CODEINE  float,
DOLOP  float,
WESTODON  float,
METHDON  float,
TALWIN  float,
ANLDKNAM  float,
ANALNONE  float,
ANALAGE  float,
ALCFIRST  float,
ALCTRY  float,
ALCREC  float,
ALCDAYS  float,
MODR30A  float,
MODR30DY  float,
UNDSTAS1  float,
VRA7AS1  float,
MRKEAAS1  float,
VRA8AS1  float,
MJKNOWN  float,
MJOPP  float,
MJFIRST  float,
MJAGE  float,
MJLIVE  float,
MJREC  float,
MJDAY30A  float,
MJTOT  float,
UNDSTAS2  float,
VRM9AS2  float,
MRKEAAS2  float,
VRM10AS2  float,
INHREAD  float,
INHOPP  float,
INHFIRST  float,
INHAGE  float,
GAS  float,
SPPAINT  float,
AEROS  float,
GLUE  float,
SOLVENT  float,
AMYLNIT  float,
ETHER  float,
NITOXID  float,
ODORIZER  float,
INHNEVER  float,
GAS30A  float,
SPPAN30A  float,
AEROS30A  float,
GLUE30A  float,
SOLVN30A  float,
AMLNT30A  float,
ETHER30A  float,
NOX30A  float,
ODR30A  float,
INH30NO  float,
INHREC  float,
INHTOT  float,
INHODRHR  float,
INHODRUS  float,
UNDSTAS3  float,
VRG10AS3  float,
MRKEAAS3  float,
VRG11AS3  float,
HALLOPP  float,
HALFIRST  float,
HALLAGE  float,
HALLREC  float,
HAL30USE  float,
HALLTOT  float,
HALPCPHR  float,
PCP  float,
HALPCP30  float,
UNDSTAS4  float,
VRL10AS4  float,
MRKEAAS4  float,
VRL11AS4  float,
COCOPP  float,
COCFIRST  float,
COCAGE  float,
COCREC  float,
COCUS30A  float,
COCTOT  float,
UNDSTAS5  float,
VRC7AS5  float,
MRKEAAS5  float,
VRC8AS5  float,
HERKNOW  float,
HEROPP  float,
HERFIRST  float,
HERAGE  float,
HERREC  float,
HER30USE  float,
HERTOT  float,
HERFRNDS  float,
HERNOADR  float,
HERNEEDL  float,
UNDSTAS6  float,
VRH11AS6  float,
MRKEAAS6  float,
VRH12AS6  float,
SPLCOC  float,
SPLHAL  float,
SPLCIG  float,
SPLHER  float,
SPLBEER  float,
SPLLQR  float,
SPLMJR  float,
SPLPILLS  float,
SPLINH  float,
GMJNOHO  float,
GMJNONE  float,
GMJMED  float,
GMJJOB  float,
GMJFUN  float,
GMJRELAX  float,
GMJAWARE  float,
GMJCNFDN  float,
GMJDEAL  float,
GMJSLEEP  float,
GMJSEX  float,
GMJAPPET  float,
GMJDK  float,
GMJMISC  float,
GMJREF1  float,
BMJCONTR  float,
BMJMEMRY  float,
BMJNONE  float,
BMJHABIT  float,
BMJSTRGR  float,
BMJHLTH  float,
BMJDIZZY  float,
BMJREFLX  float,
BMJMOOD  float,
BMJHALLU  float,
BMJAPTHY  float,
BMJJOB  float,
BMJDRIVE  float,
BMJILLEG  float,
BMJCRIME  float,
BMJEXPNS  float,
BMJDK  float,
BMJMISC  float,
BMJREF1  float,
MJHIGH  float,
MJDRHIGH  float,
MJOTHDR  float,
MJPUFFS  float,
MJDRPUFF  float,
MJOTHPUF  float,
MJINVOLV  float,
MJCAREMR  float,
MJCRMORE  float,
MJOTHMOR  float,
MJCARELS  float,
MJCRLESS  float,
MJOTHLES  float,
MJWKEND  float,
MJCRWKEN  float,
MJOTHWKN  float,
ALHIGH  float,
ALDRHIGH  float,
ALOTHDR  float,
ALSOME  float,
ALDRSOME  float,
ALOTHSOM  float,
ALOTHDRK  float,
ALYOUDRK  float,
CLOSFRNS  float,
FRNSHER  float,
FRNSEX  float,
FRNAGE  float,
FRNTRYH  float,
FRNRECH  float,
SEENUSE  float,
CONFESS  float,
TESTMNY  float,
TRACKMRK  float,
ARREST  float,
UNPRREF  float,
UNPRREP  float,
UNPRBEH  float,
UNPROTH  float,
AMBULANC  float,
DETECOTH  float,
GIVESELL  float,
TREATMNT  float,
OTHKNOW  float,
LVDHEREA  float,
LVDHEREB  float,
EVRLIVEA  float,
AGEINA1  float,
AGEOUTA1  float,
AGEINA2  float,
AGEOUTA2  float,
AGEINA3  float,
AGEOUTA3  float,
ALLLIFEA  float,
EVRLIVEB  float,
AGEINB1  float,
AGEOUTB1  float,
AGEINB2  float,
AGEOUTB2  float,
AGEINB3  float,
AGEOUTB3  float,
ALLLIFEB  float,
EVRLIVEC  float,
AGEINC1  float,
AGEOUTC1  float,
AGEINC2  float,
AGEOUTC2  float,
AGEINC3  float,
AGEOUTC3  float,
ALLLIFEC  float,
SEX  float,
RESPAGE  float,
HISPANIC  float,
HISPGRP  float,
RESPRACE  float,
RAGEGRP  float,
ENRLCOLL  float,
TYPESCHL  float,
STUDFTPT  float,
EDUC  float,
TOTPEOP  float,
UNDAGE18  float,
UNDAGE6  float,
AGE612  float,
AGE1217  float,
HHPAREN  float,
NUMPAREN  float,
HHSPOUS  float,
NUMSPOUS  float,
HHSIBLN  float,
NUMSIBLN  float,
HHOTREL  float,
NUMOTREL  float,
HHFRNDS  float,
NUMFRNDS  float,
HHOTPER  float,
NUMOTPER  float,
MARITAL  float,
EMPLOYED  float,
ROCCUP2  float,
NOLABOR  float,
CWE  float,
CWEOCC2  float,
INCOME  float,
ESTHHIN  float,
YTHSTUD  float,
YSTDFTPT  float,
YTHEDUC  float,
YTOTPEOP  float,
MOTHER  float,
FATHER  float,
OLDSIBS  float,
NUMOSIBS  float,
YNGSIBS  float,
NUMYSIBS  float,
YTHOTREL  float,
NUMYOREL  float,
YTHOTPER  float,
NUMYOPER  float,
OTHSIBS  float,
YTHEMPLD  float,
YTHOCCU2  float,
YNOLABOR  float,
HHAREA  float,
MILINSTA  float,
LOGCAMP  float,
COLLEGE  float,
RESORT  float,
CONSTR  float,
RANCH  float,
MIGRANTS  float,
TEMPRES  float,
HHTYPE  float,
UNDINT  float,
COOPINT  float,
PRIVACY  float,
ADULTYTH  float,
PAREXAMQ  float,
ADLTQCD  float,
QUEXTYPE  float,
INTVLEN  float,
FIID  float,
TOTHHVIS  float,
FINLRES1  float,
VSADLTCM  float,
PHADLTCM  float,
FINLRES2  float,
VSYTHCM  float,
PHYTHCM  float,
YTHINHH  float,
RES1825  float,
RES2649  float,
RES50OVR  float,
AGR1REL1  float,
AGR1SEX1  float,
AGR1AGE1  float,
AGR1RSP1  float,
AGR1REL2  float,
AGR1SEX2  float,
AGR1AGE2  float,
AGR1RSP2  float,
AGR1REL3  float,
AGR1SEX3  float,
AGR1AGE3  float,
AGR1RSP3  float,
AGR1REL4  float,
AGR1SEX4  float,
AGR1AGE4  float,
AGR1RSP4  float,
AGR2REL1  float,
AGR2SEX1  float,
AGR2AGE1  float,
AGR2RSP1  float,
AGR2REL2  float,
AGR2SEX2  float,
AGR2AGE2  float,
AGR2RSP2  float,
AGR2REL3  float,
AGR2SEX3  float,
AGR2AGE3  float,
AGR2RSP3  float,
AGR2REL4  float,
AGR2SEX4  float,
AGR2AGE4  float,
AGR2RSP4  float,
AGR3REL1  float,
AGR3SEX1  float,
AGR3AGE1  float,
AGR3RSP1  float,
AGR3REL2  float,
AGR3SEX2  float,
AGR3AGE2  float,
AGR3RSP2  float,
AGR3REL3  float,
AGR3SEX3  float,
AGR3AGE3  float,
AGR3RSP3  float,
AGR3REL4  float,
AGR3SEX4  float,
AGR3AGE4  float,
AGR3RSP4  float,
YTH1217  float,
YTH1REL  float,
YTH1SEX  float,
YTH1AGE  float,
YTH1RSP  float,
YTH2REL  float,
YTH2SEX  float,
YTH2AGE  float,
YTH2RSP  float,
YTH3REL  float,
YTH3SEX  float,
YTH3AGE  float,
YTH3RSP  float,
YTH4REL  float,
YTH4SEX  float,
YTH4AGE  float,
YTH4RSP  float,
REGION  float,
DIVISION  float,
POPDENX  float,
IRAGE  float,
IIAGE  float,
IRSEX  float,
IISEX  float,
IRRACEX  float,
IIRACEX  float,
IRHOIND  float,
IIHOIND  float,
IRHOGRP  float,
IIHOGRP  float,
IRMARIT  float,
IIMARIT  float,
IREDUC  float,
IIEDUC  float,
IRALCRC  float,
IIALCRC  float,
IRMJRC  float,
IIMJRC  float,
IRCOCRC  float,
IICOCRC  float,
IRSEDRC  float,
IISEDRC  float,
IRTRANRC  float,
IITRANRC  float,
IRSTIMRC  float,
IISTIMRC  float,
IRANALRC  float,
IIANALRC  float,
IRCIGRC  float,
IICIGRC  float,
IRINHRC  float,
IIINHRC  float,
IRHALLRC  float,
IIHALLRC  float,
IRHERRC  float,
IIHERRC  float,
CATAGE  float,
CATAG2  float,
CATAG3  float,
RACE  float,
HISPRACE  float,
EDUCCAT2  float,
HALFLAG  float,
HALYR  float,
HALMON  float,
STMFLAG  float,
STMYR  float,
STMMON  float,
SEDFLAG  float,
SEDYR  float,
SEDMON  float,
TRQFLAG  float,
TRQYR  float,
TRQMON  float,
ANLFLAG  float,
ANLYR  float,
ANLMON  float,
ALCFLAG  float,
ALCYR  float,
ALCMON  float,
CIGFLAG  float,
CIGYR  float,
CIGMON  float,
HERFLAG  float,
HERYR  float,
HERMON  float,
MRJFLAG  float,
MRJYR  float,
MRJMON  float,
COCFLAG  float,
COCYR  float,
COCMON  float,
INHFLAG  float,
INHYR  float,
INHMON  float,
PSYFLAG2  float,
PSYYR2  float,
PSYMON2  float,
SUMFLAG  float,
SUMYR  float,
SUMMON  float,
MJOFLAG  float,
MJOYR2  float,
MJOMON2  float,
IEMFLAG  float,
IEMYR  float,
IEMMON  float,
VESTR  float,
VEREP  float,
ANALWT float,
CANALWT float,
NANALWT float,
INITWT float,
WT1 float,
WT2 float,
CINITWT float,
CWT1 float,
CWT2 float,
NINITWT float,
NWT1 float,
NWT2 float
)

                ROW FORMAT DELIMITED
                FIELDS TERMINATED BY '	'
                LINES TERMINATED BY '\n'
                LOCATION '{drug_dir}/NHSDA-1988-DS0001-data-excel'
                TBLPROPERTIES ('skip.header.line.count'='1')
"""
```

In [129]:

```
pd.read_sql(create_table, conn)
```

Out[129]:

—

In [130]:

```
pd.read_sql(f'SELECT count(*) FROM {database_name}.{table_name2} LIMIT 5', conn
)
```

Out[130]:

	_col0
0	8814

In [126]:

```
table_name3 ='NHSDA_1995'
pd.read_sql(f'DROP TABLE IF EXISTS {database_name}.{table_name3}', conn)

create_table = f"""
CREATE EXTERNAL TABLE IF NOT EXISTS {database_name}.{table_name3}(
                CASEID  float,
RESPID  float,
ENCPSU  float,
ENCSEG  float,
ENCCASE  float,
CIGMORLS  float,
CIGTRY  float,
CIG5PK  float,
CIGREC  float,
AVCIG  float,
HRDHER  float,
HRDMJ  float,
HRDCOC  float,
HRDLSD  float,
HRDBAR  float,
HRDTRN  float,
HRDAMP  float,
ADDHER  float,
ADDALC  float,
ADDMJ  float,
ADDTOB  float,
ADDBAR  float,
ADDTRN  float,
ADDAMP  float,
ADDLSD  float,
ADDCOC  float,
ADDNONE  float,
SEDLIKE  float,
SEDFEEL  float,
SEDNEED  float,
SEDREC  float,
SED30MOA  float,
SED30MOB  float,
SED30MOC  float,
SEDDAL30  float,
BUTISOL  float,
BUTICAPS  float,
AMYTAL  float,
ESKABARB  float,
LUMINAL  float,
MEBARAL  float,
AMOBARB  float,
PHENOBAR  float,
ALURATE  float,
PLACIDYL  float,
DORIDEN  float,
NOLUDAR  float,
SOPOR  float,
QUAALUDE  float,
PAREST  float,
NOCTEC  float,
METHAQ  float,
CHHYD  float,
NEMBUTAL  float,
CARBTAL  float,
SECONAL  float,
TUINAL  float,
PENTOB  float,
SECOB  float,
DALMANE  float,
SEDDKNAM  float,
NOSEDAT  float,
SEDAGE  float,
TRNLIKE  float,
TRNFEEL  float,
TRNNEED  float,
TRANREC  float,
TRN30MOA  float,
TRN30MOB  float,
TRN30MOC  float,
TRNBEN30  float,
VALIUM  float,
LIBRIUM  float,
LIBRITAB  float,
SKLY  float,
SERAX  float,
TRANXENE  float,
ATIVAN  float,
VERSTRAN  float,
MEPRSPAN  float,
MILTOWN  float,
EQUANIL  float,
MEPROB  float,
VISTAR  float,
ATARAX  float,
BENADRYL  float,
TRDKNAM  float,
NOTRANQ  float,
TRANAGE  float,
STIMLIKE  float,
STIMFEEL  float,
STIMNEED  float,
STIMREC  float,
STM30MOA  float,
STM30MOB  float,
STMRIT30  float,
STMCYL30  float,
DEXED  float,
DEXAMYL  float,
ESKAT  float,
BENZ  float,
BIPHET  float,
DESOXYN  float,
DETAMP  float,
METHI  float,
OBLA  float,
TENUATE  float,
TEPANIL  float,
DIDREX  float,
PLEGINE  float,
PRELUDIN  float,
PRESATE  float,
IONAMIN  float,
PONDIMIN  float,
VORANIL  float,
SANOREX  float,
RITALIN  float,
CYLERT  float,
STMDKNAM  float,
NOSTIMS  float,
STIMAGE  float,
ANALLIKE  float,
ANALFEEL  float,
ANALNEED  float,
ANALREC  float,
ANL30MOA  float,
ANL30MOB  float,
ANL30MOC  float,
ANLTAL30  float,
DARVON  float,
DOLENE  float,
SK65A  float,
PROPOXY  float,
LERITINE  float,
LEVODRO  float,
PERCODAN  float,
DEMEROL  float,
DILAUD  float,
TYLCOD  float,
CODEINE  float,
DOLOP  float,
WESTODON  float,
METHDON  float,
TALWIN  float,
ANLDKNAM  float,
ANALNONE  float,
ANALAGE  float,
ALCFIRST  float,
ALCTRY  float,
ALCREC  float,
ALCDAYS  float,
MODR30A  float,
MODR30DY  float,
UNDSTAS1  float,
VRA7AS1  float,
MRKEAAS1  float,
VRA8AS1  float,
MJKNOWN  float,
MJOPP  float,
MJFIRST  float,
MJAGE  float,
MJLIVE  float,
MJREC  float,
MJDAY30A  float,
MJTOT  float,
UNDSTAS2  float,
VRM9AS2  float,
MRKEAAS2  float,
VRM10AS2  float,
INHREAD  float,
INHOPP  float,
INHFIRST  float,
INHAGE  float,
GAS  float,
SPPAINT  float,
AEROS  float,
GLUE  float,
SOLVENT  float,
AMYLNIT  float,
ETHER  float,
NITOXID  float,
ODORIZER  float,
INHNEVER  float,
GAS30A  float,
SPPAN30A  float,
AEROS30A  float,
GLUE30A  float,
SOLVN30A  float,
AMLNT30A  float,
ETHER30A  float,
NOX30A  float,
ODR30A  float,
INH30NO  float,
INHREC  float,
INHTOT  float,
INHODRHR  float,
INHODRUS  float,
UNDSTAS3  float,
VRG10AS3  float,
MRKEAAS3  float,
VRG11AS3  float,
HALLOPP  float,
HALFIRST  float,
HALLAGE  float,
HALLREC  float,
HAL30USE  float,
HALLTOT  float,
HALPCPHR  float,
PCP  float,
HALPCP30  float,
UNDSTAS4  float,
VRL10AS4  float,
MRKEAAS4  float,
VRL11AS4  float,
COCOPP  float,
COCFIRST  float,
COCAGE  float,
COCREC  float,
COCUS30A  float,
COCTOT  float,
UNDSTAS5  float,
VRC7AS5  float,
MRKEAAS5  float,
VRC8AS5  float,
HERKNOW  float,
HEROPP  float,
HERFIRST  float,
HERAGE  float,
HERREC  float,
HER30USE  float,
HERTOT  float,
HERFRNDS  float,
HERNOADR  float,
HERNEEDL  float,
UNDSTAS6  float,
VRH11AS6  float,
MRKEAAS6  float,
VRH12AS6  float,
SPLCOC  float,
SPLHAL  float,
SPLCIG  float,
SPLHER  float,
SPLBEER  float,
SPLLQR  float,
SPLMJR  float,
SPLPILLS  float,
SPLINH  float,
GMJNOHO  float,
GMJNONE  float,
GMJMED  float,
GMJJOB  float,
GMJFUN  float,
GMJRELAX  float,
GMJAWARE  float,
GMJCNFDN  float,
GMJDEAL  float,
GMJSLEEP  float,
GMJSEX  float,
GMJAPPET  float,
GMJDK  float,
GMJMISC  float,
GMJREF1  float,
BMJCONTR  float,
BMJMEMRY  float,
BMJNONE  float,
BMJHABIT  float,
BMJSTRGR  float,
BMJHLTH  float,
BMJDIZZY  float,
BMJREFLX  float,
BMJMOOD  float,
BMJHALLU  float,
BMJAPTHY  float,
BMJJOB  float,
BMJDRIVE  float,
BMJILLEG  float,
BMJCRIME  float,
BMJEXPNS  float,
BMJDK  float,
BMJMISC  float,
BMJREF1  float,
MJHIGH  float,
MJDRHIGH  float,
MJOTHDR  float,
MJPUFFS  float,
MJDRPUFF  float,
MJOTHPUF  float,
MJINVOLV  float,
MJCAREMR  float,
MJCRMORE  float,
MJOTHMOR  float,
MJCARELS  float,
MJCRLESS  float,
MJOTHLES  float,
MJWKEND  float,
MJCRWKEN  float,
MJOTHWKN  float,
ALHIGH  float,
ALDRHIGH  float,
ALOTHDR  float,
ALSOME  float,
ALDRSOME  float,
ALOTHSOM  float,
ALOTHDRK  float,
ALYOUDRK  float,
CLOSFRNS  float,
FRNSHER  float,
FRNSEX  float,
FRNAGE  float,
FRNTRYH  float,
FRNRECH  float,
SEENUSE  float,
CONFESS  float,
TESTMNY  float,
TRACKMRK  float,
ARREST  float,
UNPRREF  float,
UNPRREP  float,
UNPRBEH  float,
UNPROTH  float,
AMBULANC  float,
DETECOTH  float,
GIVESELL  float,
TREATMNT  float,
OTHKNOW  float,
LVDHEREA  float,
LVDHEREB  float,
EVRLIVEA  float,
AGEINA1  float,
AGEOUTA1  float,
AGEINA2  float,
AGEOUTA2  float,
AGEINA3  float,
AGEOUTA3  float,
ALLLIFEA  float,
EVRLIVEB  float,
AGEINB1  float,
AGEOUTB1  float,
AGEINB2  float,
AGEOUTB2  float,
AGEINB3  float,
AGEOUTB3  float,
ALLLIFEB  float,
EVRLIVEC  float,
AGEINC1  float,
AGEOUTC1  float,
AGEINC2  float,
AGEOUTC2  float,
AGEINC3  float,
AGEOUTC3  float,
ALLLIFEC  float,
SEX  float,
RESPAGE  float,
HISPANIC  float,
HISPGRP  float,
RESPRACE  float,
RAGEGRP  float,
ENRLCOLL  float,
TYPESCHL  float,
STUDFTPT  float,
EDUC  float,
TOTPEOP  float,
UNDAGE18  float,
UNDAGE6  float,
AGE612  float,
AGE1217  float,
HHPAREN  float,
NUMPAREN  float,
HHSPOUS  float,
NUMSPOUS  float,
HHSIBLN  float,
NUMSIBLN  float,
HHOTREL  float,
NUMOTREL  float,
HHFRNDS  float,
NUMFRNDS  float,
HHOTPER  float,
NUMOTPER  float,
MARITAL  float,
EMPLOYED  float,
ROCCUP2  float,
NOLABOR  float,
CWE  float,
CWEOCC2  float,
INCOME  float,
ESTHHIN  float,
YTHSTUD  float,
YSTDFTPT  float,
YTHEDUC  float,
YTOTPEOP  float,
MOTHER  float,
FATHER  float,
OLDSIBS  float,
NUMOSIBS  float,
YNGSIBS  float,
NUMYSIBS  float,
YTHOTREL  float,
NUMYOREL  float,
YTHOTPER  float,
NUMYOPER  float,
OTHSIBS  float,
YTHEMPLD  float,
YTHOCCU2  float,
YNOLABOR  float,
HHAREA  float,
MILINSTA  float,
LOGCAMP  float,
COLLEGE  float,
RESORT  float,
CONSTR  float,
RANCH  float,
MIGRANTS  float,
TEMPRES  float,
HHTYPE  float,
UNDINT  float,
COOPINT  float,
PRIVACY  float,
ADULTYTH  float,
PAREXAMQ  float,
ADLTQCD  float,
QUEXTYPE  float,
INTVLEN  float,
FIID  float,
TOTHHVIS  float,
FINLRES1  float,
VSADLTCM  float,
PHADLTCM  float,
FINLRES2  float,
VSYTHCM  float,
PHYTHCM  float,
YTHINHH  float,
RES1825  float,
RES2649  float,
RES500VR  float,
AGR1REL1  float,
AGR1SEX1  float,
AGR1AGE1  float,
AGR1RSP1  float,
AGR1REL2  float,
AGR1SEX2  float,
AGR1AGE2  float,
AGR1RSP2  float,
AGR1REL3  float,
AGR1SEX3  float,
AGR1AGE3  float,
AGR1RSP3  float,
AGR1REL4  float,
AGR1SEX4  float,
AGR1AGE4  float,
AGR1RSP4  float,
AGR2REL1  float,
AGR2SEX1  float,
AGR2AGE1  float,
AGR2RSP1  float,
AGR2REL2  float,
AGR2SEX2  float,
AGR2AGE2  float,
AGR2RSP2  float,
AGR2REL3  float,
AGR2SEX3  float,
AGR2AGE3  float,
AGR2RSP3  float,
AGR2REL4  float,
AGR2SEX4  float,
AGR2AGE4  float,
AGR2RSP4  float,
AGR3REL1  float,
AGR3SEX1  float,
AGR3AGE1  float,
AGR3RSP1  float,
AGR3REL2  float,
AGR3SEX2  float,
AGR3AGE2  float,
AGR3RSP2  float,
AGR3REL3  float,
AGR3SEX3  float,
AGR3AGE3  float,
AGR3RSP3  float,
AGR3REL4  float,
AGR3SEX4  float,
AGR3AGE4  float,
AGR3RSP4  float,
YTH1217  float,
YTH1REL  float,
YTH1SEX  float,
YTH1AGE  float,
YTH1RSP  float,
YTH2REL  float,
YTH2SEX  float,
YTH2AGE  float,
YTH2RSP  float,
YTH3REL  float,
YTH3SEX  float,
YTH3AGE  float,
YTH3RSP  float,
YTH4REL  float,
YTH4SEX  float,
YTH4AGE  float,
YTH4RSP  float,
REGION  float,
DIVISION  float,
POPDENX  float,
IRAGE  float,
IIAGE  float,
IRSEX  float,
IISEX  float,
IRRACEX  float,
IIRACEX  float,
IRHOIND  float,
IIHOIND  float,
IRHOGRP  float,
IIHOGRP  float,
IRMARIT  float,
IIMARIT  float,
IREDUC  float,
IIEDUC  float,
IRALCRC  float,
IIALCRC  float,
IRMJRC  float,
IIMJRC  float,
IRCOCRC  float,
IICOCRC  float,
IRSEDRC  float,
IISEDRC  float,
IRTRANRC  float,
IITRANRC  float,
IRSTIMRC  float,
IISTIMRC  float,
IRANALRC  float,
IIANALRC  float,
IRCIGRC  float,
IICIGRC  float,
IRINHRC  float,
IIINHRC  float,
IRHALLRC  float,
IIHALLRC  float,
IRHERRC  float,
IIHERRC  float,
CATAGE  float,
CATAG2  float,
CATAG3  float,
RACE  float,
HISPRACE  float,
EDUCCAT2  float,
HALFLAG  float,
HALYR  float,
HALMON  float,
STMFLAG  float,
STMYR  float,
STMMON  float,
SEDFLAG  float,
SEDYR  float,
SEDMON  float,
TRQFLAG  float,
TRQYR  float,
TRQMON  float,
ANLFLAG  float,
ANLYR  float,
ANLMON  float,
ALCFLAG  float,
ALCYR  float,
ALCMON  float,
CIGFLAG  float,
CIGYR  float,
CIGMON  float,
HERFLAG  float,
HERYR  float,
HERMON  float,
MRJFLAG  float,
MRJYR  float,
MRJMON  float,
COCFLAG  float,
COCYR  float,
COCMON  float,
INHFLAG  float,
INHYR  float,
INHMON  float,
PSYFLAG2  float,
PSYYR2  float,
PSYMON2  float,
SUMFLAG  float,
SUMYR  float,
SUMMON  float,
MJOFLAG  float,
MJOYR2  float,
MJOMON2  float,
IEMFLAG  float,
IEMYR  float,
IEMMON  float,
VESTR  float,
VEREP  float,
ANALWT float,
CANALWT float,
NANALWT float,
INITWT float,
WT1 float,
WT2 float,
CINITWT float,
CWT1 float,
CWT2 float,
NINITWT float,
NWT1 float,
NWT2 float
)

                ROW FORMAT DELIMITED
                FIELDS TERMINATED BY '	'
                LINES TERMINATED BY '\n'
                LOCATION '{drug_dir}/NHSDA-1995-DS0001-data-excel'
                TBLPROPERTIES ('skip.header.line.count'='1')
"""
```

In [131]:

```
pd.read_sql(create_table, conn)

pd.read_sql(f'SELECT * FROM {database_name}.{table_name3} LIMIT 2', conn)
```

Out[131]:

	caseid	respid	encpsu	encseg	enccase	cigmorls	cigtry	cig5pk	cigrec	avcig	...	nana
0	8883.0	89789.0	9404.0	2.0	59.0	705.0	9462.0	2.0	1.0	1.0	...	9
1	8884.0	89797.0	9415.0	1.0	39.0	795.0	1548.0	2.0	3.0	1.0	...	9

2 rows × 603 columns

In [132]:

```
pd.read_sql(f'SELECT * FROM {database_name}.{table_name2} LIMIT 2', conn)
```

Out[132]:

	caseid	respid	encpsu	encseg	enccase	cigmorls	cigtry	cig5pk	cigrec	avcig	...	nanalv
0	1.0	224.0	65.0	1397.0	6360.0	23.0	1.0	5.0	99.0	99.0	...	9
1	2.0	1032.0	91.0	524.0	260.0	14.0	1.0	1.0	2.0	4.0	...	9

2 rows × 603 columns

In [134]:

```
pd.read_sql(f'SELECT * FROM {database_name}.{table_name} LIMIT 2', conn)
```

Out[134]:

	caseid	respid	encpsu	encseg	enccase	cigmorls	cigtry	cig5pk	cigrec	avcig	...	nanalv
0	1.0	1214.0	63.0	151.0	2040.0	3.0	16.0	1.0	4.0	99.0	...	0
1	2.0	1478.0	63.0	151.0	5931.0	3.0	14.0	2.0	19.0	99.0	...	0

2 rows × 603 columns

In [136]:

```
pd.read_sql(f'DROP VIEW IF EXISTS all_record', conn)
```

Out[136]:

—

In [137]:

```
pd.read_sql(f'create view all_record as SELECT * FROM {database_name}.{table_na
me} union all SELECT * FROM {database_name}.{table_name2} union all SELECT * FR
OM {database_name}.{table_name3} ', conn)
```

Out[137]:

—

In [138]:

```
pd.read_sql(f'SELECT count(*) FROM all_record', conn)
```

Out[138]:

	_col0
0	33785

In [139]:

```
df = pd.read_sql(f'SELECT * FROM all_record', conn)
```

In [141]:
```
df.head()
```

Out[141]:

	caseid	respid	encpsu	encseg	enccase	cigmorls	cigtry	cig5pk	cigrec	avcig	...	nanalv
0	1.0	1214.0	63.0	151.0	2040.0	3.0	16.0	1.0	4.0	99.0	...	0
1	2.0	1478.0	63.0	151.0	5931.0	3.0	14.0	2.0	19.0	99.0	...	0
2	3.0	1608.0	63.0	151.0	4771.0	3.0	17.0	2.0	19.0	99.0	...	0
3	4.0	1661.0	63.0	151.0	292.0	2.0	91.0	91.0	91.0	91.0	...	0
4	5.0	1803.0	63.0	151.0	5232.0	2.0	15.0	2.0	19.0	99.0	...	0

5 rows × 603 columns

In [142]:
```
print('Number of Rows:', df.shape[0])
print('Number of Columns:', df.shape[1], '\n')

data_types = df.dtypes
data_types = pd.DataFrame(data_types)
data_types = data_types.assign(Null_Values =  df.isnull().sum())
data_types.reset_index(inplace = True)
data_types.rename(columns={0:'Data Type',
                           'index': 'Column/Variable',
                           'Null_Values': "# of Nulls"})
```

```
Number of Rows: 33785
Number of Columns: 603
```

Out[142]:

	Column/Variable	Data Type	# of Nulls
0	caseid	float64	0
1	respid	float64	0
2	encpsu	float64	0
3	encseg	float64	0
4	enccase	float64	0
...	...	...	...
598	cwt1	float64	0
599	cwt2	float64	0
600	ninitwt	float64	0
601	nwt1	float64	0
602	nwt2	float64	0

603 rows × 3 columns

In [143]:

```
df.corr
```

Out[143]:

```
<bound method DataFrame.corr of        caseid   respid  encpsu  encseg  enccase
cigmorls  cigtry  cig5pk  \
0         1.0   1214.0    63.0   151.0   2040.0       3.0    16.0     1.0
1         2.0   1478.0    63.0   151.0   5931.0       3.0    14.0     2.0
2         3.0   1608.0    63.0   151.0   4771.0       3.0    17.0     2.0
3         4.0   1661.0    63.0   151.0    292.0       2.0    91.0    91.0
4         5.0   1803.0    63.0   151.0   5232.0       2.0    15.0     2.0
...       ...      ...     ...     ...      ...       ...     ...     ...
33780  8878.0  89722.0  9425.0     1.0     34.0    1659.0  3225.0     2.0
33781  8879.0  89730.0  9519.0     1.0     93.0     666.0  6209.0     2.0
33782  8880.0  89748.0  9527.0     1.0     54.0    1132.0  2673.0     2.0
33783  8881.0  89755.0  9435.0     2.0     97.0     883.0  8452.0     1.0
33784  8882.0  89763.0  9433.0     1.0     66.0     314.0  7070.0     2.0

       cigrec  avcig  ...  nanalwt   initwt        wt1      wt2  cinitwt  \
0         4.0   99.0  ...      0.0   3.2720  24964.414   1.0096   3.2720
1        19.0   99.0  ...      0.0   1.5764  24964.414   1.0278   1.5764
2        19.0   99.0  ...      0.0   0.5309  24964.414   0.9780   0.5309
3        91.0   91.0  ...      0.0   2.1813  24964.414   1.0096   2.1813
4        19.0   99.0  ...      0.0   2.1813  24964.414   1.0096   2.1813
...       ...    ...  ...      ...      ...        ...      ...      ...
33780     1.0    1.0  ...     99.0  99.0000     99.000  99.0000  99.0000
33781     4.0   99.0  ...     99.0  99.0000     99.000  99.0000  99.0000
33782     4.0   99.0  ...     99.0  99.0000     99.000  99.0000  99.0000
33783    99.0   99.0  ...     99.0  99.0000     99.000  99.0000  99.0000
33784     3.0    1.0  ...     99.0  99.0000     99.000  99.0000  99.0000

            cwt1     cwt2  ninitwt       nwt1    nwt2
0      31239.281   0.9923      0.0  24964.168  1.0105
1      31239.281   1.0555      0.0  24964.168  0.8663
2      31239.281   0.9473      0.0  24964.168  1.0459
3      31239.281   0.9923      0.0  24964.168  1.0105
4      31239.281   0.9923      0.0  24964.168  1.0105
...          ...      ...      ...        ...     ...
33780     99.000  99.0000      2.0      2.000  2.0000
33781     99.000  99.0000      2.0      2.000  2.0000
33782     99.000  99.0000      2.0      2.000  2.0000
33783     99.000  99.0000      2.0      2.000  2.0000
33784     99.000  99.0000      2.0      2.000  2.0000

[33785 rows x 603 columns]>
```

In [145]:

```
print(df.shape)
```

```
(33785, 603)
```

In [146]:

```
print(df.columns)
```

```
Index(['caseid', 'respid', 'encpsu', 'encseg', 'enccase', 'cigmorls', 'cigtry',
       'cig5pk', 'cigrec', 'avcig',
       ...
       'nanalwt', 'initwt', 'wt1', 'wt2', 'cinitwt', 'cwt1', 'cwt2', 'ninitwt',
       'nwt1', 'nwt2'],
      dtype='object', length=603)
```

In [147]:
```
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 33785 entries, 0 to 33784
Columns: 603 entries, caseid to nwt2
dtypes: float64(603)
memory usage: 155.4 MB
None
```

In [148]:
```
df.describe()
```

Out[148]:

	caseid	respid	encpsu	encseg	enccase	cigmorls
count	33785.000000	33785.000000	33785.000000	33785.000000	33785.000000	33785.000000
mean	6583.728963	60815.072991	4998.708865	253.390824	1741.831760	509.685837
std	4720.702219	53405.573841	4701.591161	397.733889	2317.336228	608.026076
min	1.000000	2.000000	1.000000	1.000000	1.000000	1.000000
25%	2816.000000	9375.000000	56.000000	1.000000	61.000000	11.000000
50%	5631.000000	49692.000000	9404.000000	2.000000	109.000000	95.000000
75%	9301.000000	94029.000000	9501.000000	368.000000	3321.000000	997.000000
max	17747.000000	182295.000000	9536.000000	1532.000000	8214.000000	1908.000000

8 rows × 603 columns

In [159]:
```
#df.value_counts(normalize=True)
```

In [166]:
```
df.index
```

Out[166]:
```
RangeIndex(start=0, stop=33785, step=1)
```

In [170]:
```
import numpy as np
```

In [172]:
```
import matplotlib.pyplot as plt
```

In [ ]:
```
filtered_df=df.iloc[0:33785, 0:603]
filtered_df.apply(np.max)
```

```
In [179]: df.pivot_table(['cigmorls', 'cigtry', 'cig5pk', 'cigrec'],
                         ['nanalwt'], aggfunc='mean')
```

```
Out[179]:
                cig5pk     cigmorls     cigrec        cigtry
  nanalwt
     0.00    17.326694     2.366286  21.949478     28.089871
     1.00     9.064935   534.123377  26.155844   4708.285714
     2.00     5.961832   716.910305  25.171756   6000.339695
     3.00    14.358025    66.049383  46.086420    261.827160
     4.00    19.920455    28.875000  63.704545     19.784091
      ...          ...          ...        ...           ...
772608.20     1.000000     2.000000   4.000000     15.000000
789707.50     1.000000     2.000000   1.000000     15.000000
809619.40     2.000000     1.000000  19.000000     17.000000
874388.80    91.000000     2.000000  91.000000     91.000000
961537.44     1.000000     2.000000   1.000000     16.000000

2078 rows × 4 columns
```

```
In [ ]:
```