ai minor project for car price prediction

November 27, 2024

[ ]:
```
 # Load the 2024 dataset
dataset = pd.read_csv("dataset.csv")

# Initial exploration
dataset.head()
```

[ ]:
```
   Unnamed: 0                                Name    Location  Year  \
0           0            Maruti Wagon R LXI CNG      Mumbai  2010
1           1  Hyundai Creta 1.6 CRDi SX Option        Pune  2015
2           2                      Honda Jazz V     Chennai  2011
3           3                 Maruti Ertiga VDI     Chennai  2012
4           4   Audi A4 New 2.0 TDI Multitronic  Coimbatore  2013

   Kilometers_Driven Fuel_Type Transmission Owner_Type     Mileage   Engine  \
0              72000       CNG       Manual      First  26.6 km/kg   998 CC
1              41000    Diesel       Manual      First  19.67 kmpl  1582 CC
2              46000    Petrol       Manual      First   18.2 kmpl  1199 CC
3              87000    Diesel       Manual      First  20.77 kmpl  1248 CC
4              40670    Diesel    Automatic     Second   15.2 kmpl  1968 CC

       Power  Seats  New_Price  Price
0  58.16 bhp    5.0        NaN   1.75
1  126.2 bhp    5.0        NaN  12.50
2   88.7 bhp    5.0  8.61 Lakh   4.50
3  88.76 bhp    7.0        NaN   6.00
4  140.8 bhp    5.0        NaN  17.74
```

[ ]:
```
import datetime
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score
```

[2]:
```
from google.colab import drive
drive.mount('/content/drive')
!pip install nbconvert
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call
drive.mount("/content/drive", force_remount=True).
Requirement already satisfied: nbconvert in /usr/local/lib/python3.10/dist-
packages (7.16.4)
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-
packages (from nbconvert) (4.12.3)
Requirement already satisfied: bleach!=5.0.0 in /usr/local/lib/python3.10/dist-
packages (from nbconvert) (6.2.0)
Requirement already satisfied: defusedxml in /usr/local/lib/python3.10/dist-
packages (from nbconvert) (0.7.1)
Requirement already satisfied: jinja2>=3.0 in /usr/local/lib/python3.10/dist-
packages (from nbconvert) (3.1.4)
Requirement already satisfied: jupyter-core>=4.7 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (5.7.2)
Requirement already satisfied: jupyterlab-pygments in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (0.3.0)
Requirement already satisfied: markupsafe>=2.0 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (3.0.2)
Requirement already satisfied: mistune<4,>=2.0.3 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (3.0.2)
Requirement already satisfied: nbclient>=0.5.0 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (0.10.0)
Requirement already satisfied: nbformat>=5.7 in /usr/local/lib/python3.10/dist-
packages (from nbconvert) (5.10.4)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-
packages (from nbconvert) (24.2)
Requirement already satisfied: pandocfilters>=1.4.1 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (1.5.1)
Requirement already satisfied: pygments>=2.4.1 in
/usr/local/lib/python3.10/dist-packages (from nbconvert) (2.18.0)
Requirement already satisfied: tinycss2 in /usr/local/lib/python3.10/dist-
packages (from nbconvert) (1.4.0)
Requirement already satisfied: traitlets>=5.1 in /usr/local/lib/python3.10/dist-
packages (from nbconvert) (5.7.1)
Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-
packages (from bleach!=5.0.0->nbconvert) (0.5.1)
Requirement already satisfied: platformdirs>=2.5 in
/usr/local/lib/python3.10/dist-packages (from jupyter-core>=4.7->nbconvert)
(4.3.6)
Requirement already satisfied: jupyter-client>=6.1.12 in
/usr/local/lib/python3.10/dist-packages (from nbclient>=0.5.0->nbconvert)
(6.1.12)
Requirement already satisfied: fastjsonschema>=2.15 in
```

```
/usr/local/lib/python3.10/dist-packages (from nbformat>=5.7->nbconvert) (2.20.0)
Requirement already satisfied: jsonschema>=2.6 in
/usr/local/lib/python3.10/dist-packages (from nbformat>=5.7->nbconvert) (4.23.0)
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-
packages (from beautifulsoup4->nbconvert) (2.6)
Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.10/dist-
packages (from jsonschema>=2.6->nbformat>=5.7->nbconvert) (24.2.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in
/usr/local/lib/python3.10/dist-packages (from
jsonschema>=2.6->nbformat>=5.7->nbconvert) (2024.10.1)
Requirement already satisfied: referencing>=0.28.4 in
/usr/local/lib/python3.10/dist-packages (from
jsonschema>=2.6->nbformat>=5.7->nbconvert) (0.35.1)
Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.10/dist-
packages (from jsonschema>=2.6->nbformat>=5.7->nbconvert) (0.21.0)
Requirement already satisfied: pyzmq>=13 in /usr/local/lib/python3.10/dist-
packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert) (24.0.1)
Requirement already satisfied: python-dateutil>=2.1 in
/usr/local/lib/python3.10/dist-packages (from jupyter-
client>=6.1.12->nbclient>=0.5.0->nbconvert) (2.8.2)
Requirement already satisfied: tornado>=4.1 in /usr/local/lib/python3.10/dist-
packages (from jupyter-client>=6.1.12->nbclient>=0.5.0->nbconvert) (6.3.3)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-
packages (from python-dateutil>=2.1->jupyter-
client>=6.1.12->nbclient>=0.5.0->nbconvert) (1.16.0)
```

[ ]:
```
!apt-get install pandoc
!apt-get install texlive-xetex

!jupyter nbconvert --to pdf '/content/drive/My Drive/Colab Notebooks/ai minor␣
↪project for car price prediction.ipynb'
```

```
Reading package lists… Done
Building dependency tree… Done
Reading state information… Done
pandoc is already the newest version (2.9.2.1-3ubuntu2).
0 upgraded, 0 newly installed, 0 to remove and 49 not upgraded.
Reading package lists… Done
Building dependency tree… Done
Reading state information… Done
The following additional packages will be installed:
  dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono fonts-
texgyre
  fonts-urw-base35 libapache-pom-java libcommons-logging-java libcommons-parent-
java
  libfontbox-java libfontenc1 libgs9 libgs9-common libidn12 libijs-0.35
libjbig2dec0 libkpathsea6
  libpdfbox-java libptexenc1 libruby3.0 libsynctex2 libteckit0 libtexlua53
```

```
libtexluajit2 libwoff1
  libzzip-0-13 lmodern poppler-data preview-latex-style rake ruby ruby-net-
telnet ruby-rubygems
  ruby-webrick ruby-xmlrpc ruby3.0 rubygems-integration t1utils teckit tex-
common tex-gyre
  texlive-base texlive-binaries texlive-fonts-recommended texlive-latex-base
texlive-latex-extra
  texlive-latex-recommended texlive-pictures texlive-plain-generic tipa xfonts-
encodings
  xfonts-utils
Suggested packages:
  fonts-noto fonts-freefont-otf | fonts-freefont-ttf libavalon-framework-java
  libcommons-logging-java-doc libexcalibur-logkit-java liblog4j1.2-java poppler-
utils ghostscript
  fonts-japanese-mincho | fonts-ipafont-mincho fonts-japanese-gothic | fonts-
ipafont-gothic
  fonts-arphic-ukai fonts-arphic-uming fonts-nanum ri ruby-dev bundler debhelper
gv
  | postscript-viewer perl-tk xpdf | pdf-viewer xzdec texlive-fonts-recommended-
doc
  texlive-latex-base-doc python3-pygments icc-profiles libfile-which-perl
  libspreadsheet-parseexcel-perl texlive-latex-extra-doc texlive-latex-
recommended-doc
  texlive-luatex texlive-pstricks dot2tex prerex texlive-pictures-doc vprerex
default-jre-headless
  tipa-doc
The following NEW packages will be installed:
  dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono fonts-
texgyre
  fonts-urw-base35 libapache-pom-java libcommons-logging-java libcommons-parent-
java
  libfontbox-java libfontenc1 libgs9 libgs9-common libidn12 libijs-0.35
libjbig2dec0 libkpathsea6
  libpdfbox-java libptexenc1 libruby3.0 libsynctex2 libteckit0 libtexlua53
libtexluajit2 libwoff1
  libzzip-0-13 lmodern poppler-data preview-latex-style rake ruby ruby-net-
telnet ruby-rubygems
  ruby-webrick ruby-xmlrpc ruby3.0 rubygems-integration t1utils teckit tex-
common tex-gyre
  texlive-base texlive-binaries texlive-fonts-recommended texlive-latex-base
texlive-latex-extra
  texlive-latex-recommended texlive-pictures texlive-plain-generic texlive-xetex
tipa
  xfonts-encodings xfonts-utils
0 upgraded, 54 newly installed, 0 to remove and 49 not upgraded.
Need to get 182 MB of archives.
After this operation, 571 MB of additional disk space will be used.
Get:1 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-droid-fallback all
```

```
1:6.0.1r16-1.1build1 [1,805 kB]
Get:2 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-lato all 2.0-2.1
[2,696 kB]
Get:3 http://archive.ubuntu.com/ubuntu jammy/main amd64 poppler-data all
0.4.11-1 [2,171 kB]
Get:4 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-common all 6.17
[33.7 kB]
Get:5 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-urw-base35 all
20200910-1 [6,367 kB]
Get:6 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libgs9-common
all 9.55.0~dfsg1-0ubuntu5.10 [752 kB]
Get:7 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libidn12 amd64
1.38-4ubuntu1 [60.0 kB]
Get:8 http://archive.ubuntu.com/ubuntu jammy/main amd64 libijs-0.35 amd64
0.35-15build2 [16.5 kB]
Get:9 http://archive.ubuntu.com/ubuntu jammy/main amd64 libjbig2dec0 amd64
0.19-3build2 [64.7 kB]
Get:10 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libgs9 amd64
9.55.0~dfsg1-0ubuntu5.10 [5,031 kB]
Get:11 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libkpathsea6
amd64 2021.20210626.59705-1ubuntu0.2 [60.4 kB]
Get:12 http://archive.ubuntu.com/ubuntu jammy/main amd64 libwoff1 amd64
1.0.2-1build4 [45.2 kB]
Get:13 http://archive.ubuntu.com/ubuntu jammy/universe amd64 dvisvgm amd64
2.13.1-1 [1,221 kB]
Get:14 http://archive.ubuntu.com/ubuntu jammy/universe amd64 fonts-lmodern all
2.004.5-6.1 [4,532 kB]
Get:15 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-noto-mono all
20201225-1build1 [397 kB]
Get:16 http://archive.ubuntu.com/ubuntu jammy/universe amd64 fonts-texgyre all
20180621-3.1 [10.2 MB]
Get:17 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libapache-pom-java
all 18-1 [4,720 B]
Get:18 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libcommons-parent-
java all 43-1 [10.8 kB]
Get:19 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libcommons-logging-
java all 1.2-2 [60.3 kB]
Get:20 http://archive.ubuntu.com/ubuntu jammy/main amd64 libfontenc1 amd64
1:1.1.4-1build3 [14.7 kB]
Get:21 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libptexenc1
amd64 2021.20210626.59705-1ubuntu0.2 [39.1 kB]
Get:22 http://archive.ubuntu.com/ubuntu jammy/main amd64 rubygems-integration
all 1.18 [5,336 B]
Get:23 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 ruby3.0 amd64
3.0.2-7ubuntu2.8 [50.1 kB]
Get:24 http://archive.ubuntu.com/ubuntu jammy/main amd64 ruby-rubygems all
3.3.5-2 [228 kB]
Get:25 http://archive.ubuntu.com/ubuntu jammy/main amd64 ruby amd64 1:3.0~exp1
```

```
[5,100 B]
Get:26 http://archive.ubuntu.com/ubuntu jammy/main amd64 rake all 13.0.6-2 [61.7
kB]
Get:27 http://archive.ubuntu.com/ubuntu jammy/main amd64 ruby-net-telnet all
0.1.1-2 [12.6 kB]
Get:28 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 ruby-webrick
all 1.7.0-3ubuntu0.1 [52.1 kB]
Get:29 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 ruby-xmlrpc all
0.3.2-1ubuntu0.1 [24.9 kB]
Get:30 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libruby3.0
amd64 3.0.2-7ubuntu2.8 [5,113 kB]
Get:31 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libsynctex2
amd64 2021.20210626.59705-1ubuntu0.2 [55.6 kB]
Get:32 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libteckit0 amd64
2.5.11+ds1-1 [421 kB]
Get:33 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libtexlua53
amd64 2021.20210626.59705-1ubuntu0.2 [120 kB]
Get:34 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libtexluajit2
amd64 2021.20210626.59705-1ubuntu0.2 [267 kB]
Get:35 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libzzip-0-13 amd64
0.13.72+dfsg.1-1.1 [27.0 kB]
Get:36 http://archive.ubuntu.com/ubuntu jammy/main amd64 xfonts-encodings all
1:1.0.5-0ubuntu2 [578 kB]
Get:37 http://archive.ubuntu.com/ubuntu jammy/main amd64 xfonts-utils amd64
1:7.7+6build2 [94.6 kB]
Get:38 http://archive.ubuntu.com/ubuntu jammy/universe amd64 lmodern all
2.004.5-6.1 [9,471 kB]
Get:39 http://archive.ubuntu.com/ubuntu jammy/universe amd64 preview-latex-style
all 12.2-1ubuntu1 [185 kB]
Get:40 http://archive.ubuntu.com/ubuntu jammy/main amd64 t1utils amd64
1.41-4build2 [61.3 kB]
Get:41 http://archive.ubuntu.com/ubuntu jammy/universe amd64 teckit amd64
2.5.11+ds1-1 [699 kB]
Get:42 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-gyre all
20180621-3.1 [6,209 kB]
Get:43 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 texlive-
binaries amd64 2021.20210626.59705-1ubuntu0.2 [9,860 kB]
Get:44 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-base all
2021.20220204-1 [21.0 MB]
Get:45 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-fonts-
recommended all 2021.20220204-1 [4,972 kB]
Get:46 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-base
all 2021.20220204-1 [1,128 kB]
Get:47 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libfontbox-java all
1:1.8.16-2 [207 kB]
Get:48 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libpdfbox-java all
1:1.8.16-2 [5,199 kB]
Get:49 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-
```

```
recommended all 2021.20220204-1 [14.4 MB]
Get:50 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-pictures
all 2021.20220204-1 [8,720 kB]
Get:51 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-extra
all 2021.20220204-1 [13.9 MB]
Get:52 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-plain-
generic all 2021.20220204-1 [27.5 MB]
Get:53 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tipa all 2:1.3-21
[2,967 kB]
Get:54 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-xetex all
2021.20220204-1 [12.4 MB]
Fetched 182 MB in 8s (23.2 MB/s)
Extracting templates from packages: 100%
Preconfiguring packages …
Selecting previously unselected package fonts-droid-fallback.
(Reading database … 123855 files and directories currently installed.)
Preparing to unpack …/00-fonts-droid-fallback_1%3a6.0.1r16-1.1build1_all.deb
…
Unpacking fonts-droid-fallback (1:6.0.1r16-1.1build1) …
Selecting previously unselected package fonts-lato.
Preparing to unpack …/01-fonts-lato_2.0-2.1_all.deb …
Unpacking fonts-lato (2.0-2.1) …
Selecting previously unselected package poppler-data.
Preparing to unpack …/02-poppler-data_0.4.11-1_all.deb …
Unpacking poppler-data (0.4.11-1) …
Selecting previously unselected package tex-common.
Preparing to unpack …/03-tex-common_6.17_all.deb …
Unpacking tex-common (6.17) …
Selecting previously unselected package fonts-urw-base35.
Preparing to unpack …/04-fonts-urw-base35_20200910-1_all.deb …
Unpacking fonts-urw-base35 (20200910-1) …
Selecting previously unselected package libgs9-common.
Preparing to unpack …/05-libgs9-common_9.55.0~dfsg1-0ubuntu5.10_all.deb …
Unpacking libgs9-common (9.55.0~dfsg1-0ubuntu5.10) …
Selecting previously unselected package libidn12:amd64.
Preparing to unpack …/06-libidn12_1.38-4ubuntu1_amd64.deb …
Unpacking libidn12:amd64 (1.38-4ubuntu1) …
Selecting previously unselected package libijs-0.35:amd64.
Preparing to unpack …/07-libijs-0.35_0.35-15build2_amd64.deb …
Unpacking libijs-0.35:amd64 (0.35-15build2) …
Selecting previously unselected package libjbig2dec0:amd64.
Preparing to unpack …/08-libjbig2dec0_0.19-3build2_amd64.deb …
Unpacking libjbig2dec0:amd64 (0.19-3build2) …
Selecting previously unselected package libgs9:amd64.
Preparing to unpack …/09-libgs9_9.55.0~dfsg1-0ubuntu5.10_amd64.deb …
Unpacking libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.10) …
Selecting previously unselected package libkpathsea6:amd64.
Preparing to unpack …/10-libkpathsea6_2021.20210626.59705-1ubuntu0.2_amd64.deb
```

```
…
Unpacking libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.2) …
Selecting previously unselected package libwoff1:amd64.
Preparing to unpack …/11-libwoff1_1.0.2-1build4_amd64.deb …
Unpacking libwoff1:amd64 (1.0.2-1build4) …
Selecting previously unselected package dvisvgm.
Preparing to unpack …/12-dvisvgm_2.13.1-1_amd64.deb …
Unpacking dvisvgm (2.13.1-1) …
Selecting previously unselected package fonts-lmodern.
Preparing to unpack …/13-fonts-lmodern_2.004.5-6.1_all.deb …
Unpacking fonts-lmodern (2.004.5-6.1) …
Selecting previously unselected package fonts-noto-mono.
Preparing to unpack …/14-fonts-noto-mono_20201225-1build1_all.deb …
Unpacking fonts-noto-mono (20201225-1build1) …
Selecting previously unselected package fonts-texgyre.
Preparing to unpack …/15-fonts-texgyre_20180621-3.1_all.deb …
Unpacking fonts-texgyre (20180621-3.1) …
Selecting previously unselected package libapache-pom-java.
Preparing to unpack …/16-libapache-pom-java_18-1_all.deb …
Unpacking libapache-pom-java (18-1) …
Selecting previously unselected package libcommons-parent-java.
Preparing to unpack …/17-libcommons-parent-java_43-1_all.deb …
Unpacking libcommons-parent-java (43-1) …
Selecting previously unselected package libcommons-logging-java.
Preparing to unpack …/18-libcommons-logging-java_1.2-2_all.deb …
Unpacking libcommons-logging-java (1.2-2) …
Selecting previously unselected package libfontenc1:amd64.
Preparing to unpack …/19-libfontenc1_1%3a1.1.4-1build3_amd64.deb …
Unpacking libfontenc1:amd64 (1:1.1.4-1build3) …
Selecting previously unselected package libptexenc1:amd64.
Preparing to unpack …/20-libptexenc1_2021.20210626.59705-1ubuntu0.2_amd64.deb
…
Unpacking libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.2) …
Selecting previously unselected package rubygems-integration.
Preparing to unpack …/21-rubygems-integration_1.18_all.deb …
Unpacking rubygems-integration (1.18) …
Selecting previously unselected package ruby3.0.
Preparing to unpack …/22-ruby3.0_3.0.2-7ubuntu2.8_amd64.deb …
Unpacking ruby3.0 (3.0.2-7ubuntu2.8) …
Selecting previously unselected package ruby-rubygems.
Preparing to unpack …/23-ruby-rubygems_3.3.5-2_all.deb …
Unpacking ruby-rubygems (3.3.5-2) …
Selecting previously unselected package ruby.
Preparing to unpack …/24-ruby_1%3a3.0~exp1_amd64.deb …
Unpacking ruby (1:3.0~exp1) …
Selecting previously unselected package rake.
Preparing to unpack …/25-rake_13.0.6-2_all.deb …
Unpacking rake (13.0.6-2) …
```

```
Selecting previously unselected package ruby-net-telnet.
Preparing to unpack …/26-ruby-net-telnet_0.1.1-2_all.deb …
Unpacking ruby-net-telnet (0.1.1-2) …
Selecting previously unselected package ruby-webrick.
Preparing to unpack …/27-ruby-webrick_1.7.0-3ubuntu0.1_all.deb …
Unpacking ruby-webrick (1.7.0-3ubuntu0.1) …
Selecting previously unselected package ruby-xmlrpc.
Preparing to unpack …/28-ruby-xmlrpc_0.3.2-1ubuntu0.1_all.deb …
Unpacking ruby-xmlrpc (0.3.2-1ubuntu0.1) …
Selecting previously unselected package libruby3.0:amd64.
Preparing to unpack …/29-libruby3.0_3.0.2-7ubuntu2.8_amd64.deb …
Unpacking libruby3.0:amd64 (3.0.2-7ubuntu2.8) …
Selecting previously unselected package libsynctex2:amd64.
Preparing to unpack …/30-libsynctex2_2021.20210626.59705-1ubuntu0.2_amd64.deb
…
Unpacking libsynctex2:amd64 (2021.20210626.59705-1ubuntu0.2) …
Selecting previously unselected package libteckit0:amd64.
Preparing to unpack …/31-libteckit0_2.5.11+ds1-1_amd64.deb …
Unpacking libteckit0:amd64 (2.5.11+ds1-1) …
Selecting previously unselected package libtexlua53:amd64.
Preparing to unpack …/32-libtexlua53_2021.20210626.59705-1ubuntu0.2_amd64.deb
…
Unpacking libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.2) …
Selecting previously unselected package libtexluajit2:amd64.
Preparing to unpack
…/33-libtexluajit2_2021.20210626.59705-1ubuntu0.2_amd64.deb …
Unpacking libtexluajit2:amd64 (2021.20210626.59705-1ubuntu0.2) …
Selecting previously unselected package libzzip-0-13:amd64.
Preparing to unpack …/34-libzzip-0-13_0.13.72+dfsg.1-1.1_amd64.deb …
Unpacking libzzip-0-13:amd64 (0.13.72+dfsg.1-1.1) …
Selecting previously unselected package xfonts-encodings.
Preparing to unpack …/35-xfonts-encodings_1%3a1.0.5-0ubuntu2_all.deb …
Unpacking xfonts-encodings (1:1.0.5-0ubuntu2) …
Selecting previously unselected package xfonts-utils.
Preparing to unpack …/36-xfonts-utils_1%3a7.7+6build2_amd64.deb …
Unpacking xfonts-utils (1:7.7+6build2) …
Selecting previously unselected package lmodern.
Preparing to unpack …/37-lmodern_2.004.5-6.1_all.deb …
Unpacking lmodern (2.004.5-6.1) …
Selecting previously unselected package preview-latex-style.
Preparing to unpack …/38-preview-latex-style_12.2-1ubuntu1_all.deb …
Unpacking preview-latex-style (12.2-1ubuntu1) …
Selecting previously unselected package t1utils.
Preparing to unpack …/39-t1utils_1.41-4build2_amd64.deb …
Unpacking t1utils (1.41-4build2) …
Selecting previously unselected package teckit.
Preparing to unpack …/40-teckit_2.5.11+ds1-1_amd64.deb …
Unpacking teckit (2.5.11+ds1-1) …
```

```
Selecting previously unselected package tex-gyre.
Preparing to unpack …/41-tex-gyre_20180621-3.1_all.deb …
Unpacking tex-gyre (20180621-3.1) …
Selecting previously unselected package texlive-binaries.
Preparing to unpack …/42-texlive-
binaries_2021.20210626.59705-1ubuntu0.2_amd64.deb …
Unpacking texlive-binaries (2021.20210626.59705-1ubuntu0.2) …
Selecting previously unselected package texlive-base.
Preparing to unpack …/43-texlive-base_2021.20220204-1_all.deb …
Unpacking texlive-base (2021.20220204-1) …
```

[ ]:
```
# Data Preprocessing: Cleaning steps (dropping columns, filling missing values,␣
 ↪converting types) as done previously
X_train = X_train.iloc[:, 1:]  # Removing 'Unnamed: 0' index column
X_test = X_test.iloc[:, 1:]
```

[ ]:
```
# Splitting into train and test datasets
X_train, X_test, y_train, y_test = train_test_split(dataset.iloc[:, :-1],
                                                    dataset.iloc[:, -1],
                                                    test_size = 0.3,
                                                    random_state = 42)
```

[ ]:
```
# Extract manufacturer from 'Name' column
make_train = X_train["Name"].str.split(" ", expand = True)
make_test = X_test["Name"].str.split(" ", expand = True)
X_train["Manufacturer"] = make_train[0]
X_test["Manufacturer"] = make_test[0]
X_train.drop("Name", axis = 1, inplace = True)
X_test.drop("Name", axis = 1, inplace = True)
```

[ ]:
```
# Removing the 'Location' column as it is irrelevant
X_train.drop("Location", axis = 1, inplace = True)
X_test.drop("Location", axis = 1, inplace = True)
```

[ ]:
```
# Handling 'Year' column to represent the car's age
curr_time = datetime.datetime.now()
X_train['Year'] = X_train['Year'].apply(lambda x : curr_time.year - x)
X_test['Year'] = X_test['Year'].apply(lambda x : curr_time.year - x)
```

[ ]:
```
# Handle categorical columns (Fuel_Type, Transmission, Owner_Type)
X_train = pd.get_dummies(X_train, columns=["Manufacturer", "Fuel_Type",␣
 ↪"Transmission", "Owner_Type"], drop_first=True)
X_test = pd.get_dummies(X_test, columns=["Manufacturer", "Fuel_Type",␣
 ↪"Transmission", "Owner_Type"], drop_first=True)
```

[ ]:
```
# Handling missing columns between train and test set
missing_cols = set(X_train.columns) - set(X_test.columns)
```

```
for col in missing_cols:
    X_test[col] = 0
X_test = X_test[X_train.columns]
```

[ ]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

# Load dataset
df = pd.read_csv('dataset.csv')

# Inspect the dataset for anomalies
print(df.head())  # Ensure columns are correctly aligned and data types are␣
 ↪valid

# Define a function to clean and extract Mileage values
def clean_Mileage(value):
    """
    Cleans and extracts numeric Mileage values from strings.
    Handles cases where Mileage is specified in 'kmpl' or 'km/kg'.
    Returns None for invalid or non-numeric entries.
    """
    try:
        if isinstance(value, str):  # Check if value is a string
            if 'kmpl' in value:
                return float(value.replace(' kmpl', ''))
            elif 'km/kg' in value:
                return float(value.replace(' km/kg', ''))
        return None  # Return None for non-matching or invalid values
    except Exception as e:
        print(f"Error processing value '{value}': {e}")
        return None

# Apply the cleaning function to the Mileage column
if 'Mileage' in df.columns:
    df['Mileage'] = df['Mileage'].apply(clean_Mileage)
else:
    raise ValueError("Mileage column not found in the dataset.")

# Drop rows with invalid or missing Mileage values
df = df.dropna(subset=['Mileage'])

# Ensure the target column is present
target_column = 'Price'  # Replace with your actual target column name
if target_column not in df.columns:
    raise ValueError(f"Target column '{target_column}' not found in the dataset.
 ↪ Check your dataset.")
```

```
# Separate features and target variable
X = df.drop(target_column, axis=1)  # Drop target column to isolate features
y = df[target_column]  # Target variable

# Handle categorical features (convert to numeric using one-hot encoding)
categorical_columns = X.select_dtypes(include=['object']).columns
if not categorical_columns.empty:
    X = pd.get_dummies(X, columns=categorical_columns, drop_first=True)

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=42)

# Scale the feature data
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)

# Confirm successful processing
print("Data preprocessing completed successfully!")
print(f"Training data shape: {X_train.shape}")
print(f"Testing data shape: {X_test.shape}")
```

```
   Unnamed: 0                                Name    Location  Year  \
0           0             Maruti Wagon R LXI CNG      Mumbai  2010
1           1  Hyundai Creta 1.6 CRDi SX Option        Pune  2015
2           2                      Honda Jazz V     Chennai  2011
3           3                 Maruti Ertiga VDI     Chennai  2012
4           4   Audi A4 New 2.0 TDI Multitronic  Coimbatore  2013

   Kilometers_Driven Fuel_Type Transmission Owner_Type     Mileage   Engine  \
0              72000       CNG       Manual      First  26.6 km/kg   998 CC
1              41000    Diesel       Manual      First  19.67 kmpl  1582 CC
2              46000    Petrol       Manual      First   18.2 kmpl  1199 CC
3              87000    Diesel       Manual      First  20.77 kmpl  1248 CC
4              40670    Diesel    Automatic     Second   15.2 kmpl  1968 CC

       Power  Seats  New_Price  Price
0  58.16 bhp    5.0        NaN   1.75
1  126.2 bhp    5.0        NaN  12.50
2   88.7 bhp    5.0  8.61 Lakh   4.50
3  88.76 bhp    7.0        NaN   6.00
4  140.8 bhp    5.0        NaN  17.74
Data preprocessing completed successfully!
Training data shape: (4813, 2948)
Testing data shape: (1204, 2948)
```

1 training model and evaluation

[ ]:
```
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import r2_score
from sklearn.impute import SimpleImputer

# Impute missing values in features
imputer = SimpleImputer(strategy='mean')  # Use mean to replace NaN values
X_train = imputer.fit_transform(X_train)
X_test = imputer.transform(X_test)

# Linear Regression Model
linearRegression = LinearRegression()
linearRegression.fit(X_train, y_train)
y_pred_lr = linearRegression.predict(X_test)
r2_lr = r2_score(y_test, y_pred_lr)
print(f"Linear Regression R2 score: {r2_lr}")

# Random Forest Regressor Model
rf = RandomForestRegressor(n_estimators=100)
rf.fit(X_train, y_train)
y_pred_rf = rf.predict(X_test)
r2_rf = r2_score(y_test, y_pred_rf)
print(f"Random Forest R2 score: {r2_rf}")
```

```
Linear Regression R2 score: -9.949818030280959e+26
Random Forest R2 score: 0.8267315920293095
```

2 model accurracy

[ ]:
```
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestRegressor
import joblib
from sklearn.model_selection import train_test_split

# Load the dataset
dataset = pd.read_csv("car predict 2025 assum.csv")  # Replace with the correct␣
 ↪path

# Check for non-numeric data and inspect columns
print(dataset.head())  # View the first few rows to check the data
```

```
# Remove any non-numeric columns that aren't useful for prediction (e.g.,␣
 ↪'Brand', 'Model')
dataset.drop(['Brand', 'Model'], axis=1, inplace=True)  # Drop 'Brand' and␣
 ↪'Model' columns

# Apply necessary preprocessing steps
dataset['Year'] = dataset['Year'].apply(lambda x: 2025 - x)  # Adjusting the␣
 ↪year for prediction

# Handle Fuel_Type by encoding it (it's a categorical feature)
make_train = dataset["Fuel_Type"].str.split(" ", expand=True)
dataset["Fuel_Type"] = make_train[0]  # Simplify or handle accordingly

# Apply one-hot encoding for categorical columns like 'Fuel_Type' and␣
 ↪'Transmission'
dataset = pd.get_dummies(dataset, columns=["Fuel_Type", "Transmission"],␣
 ↪drop_first=True)

# Check again for non-numeric columns after transformation
print(dataset.head())  # View the dataset after encoding

# Define features (X) and target (y)
X = dataset.drop("Price", axis=1)  # All columns except 'Price'
y = dataset['Price']  # The target variable is 'Price'

# Ensure all columns are numeric
print(X.dtypes)  # Ensure no non-numeric columns are present

# Convert any remaining non-numeric columns if they exist (e.g.,␣
 ↪'Fuel_Type_Petrol' or 'Transmission_Manual')
# This is usually handled by pd.get_dummies(), but if there are any unexpected␣
 ↪string columns, handle them

# Split data into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
 ↪random_state=42)

# Initialize the RandomForest model
rf = RandomForestRegressor(n_estimators=100, random_state=42)

# Fit the model
rf.fit(X_train, y_train)

# Save the trained Random Forest model
joblib.dump(rf, 'rf_model.pkl')
```

```
# Save the StandardScaler used for scaling (if necessary)
standardScaler = StandardScaler()
X_train_scaled = standardScaler.fit_transform(X_train)
joblib.dump(standardScaler, 'scaler.pkl')

# Evaluate the model (optional)
print("Model accuracy on test set:", rf.score(X_test, y_test))
```

```
    Brand     Model  Year  Mileage Fuel_Type Transmission  Price
0  Toyota   Corolla  2025     15.5    Petrol    Automatic  25000
1   Honda     Civic  2025     12.0    Diesel       Manual  28000
2    Ford     Focus  2025     13.0    Petrol       Manual  23000
3     BMW  3 Series  2025     10.5    Diesel    Automatic  40000
4    Audi        A4  2025     11.5    Petrol    Automatic  35000
   Year  Mileage  Price  Fuel_Type_Petrol  Transmission_Manual
0     0     15.5  25000              True                False
1     0     12.0  28000             False                 True
2     0     13.0  23000              True                 True
3     0     10.5  40000             False                False
4     0     11.5  35000              True                False
Year                     int64
Mileage                float64
Fuel_Type_Petrol          bool
Transmission_Manual       bool
dtype: object
Model accuracy on test set: -9.803025
```

3 convert to numeric value

[ ]:
```
import pandas as pd
from sklearn.preprocessing import StandardScaler

# Load the dataset from CSV file
df = pd.read_csv('dataset.csv')

# Clean the 'Mileage' column: remove non-numeric characters (e.g., ' km/kg')
df['Mileage'] = df['Mileage'].replace(r'\D', '', regex=True).astype(float)

# Now you can apply StandardScaler
scaler = StandardScaler()

# Fit the scaler on the 'Mileage' data
scaler.fit(df[['Mileage']])

# If you want to scale the dataset for prediction, for example, for X_2024
X_2024 = df[['Mileage']]  # Replace with your actual dataset columns
X_2024_scaled = scaler.transform(X_2024)
```

```
# Print the scaled data
print(X_2024_scaled)
```

```
[[-0.68556048]
 [ 1.29098365]
 [-0.78316759]
 …
 [-0.83197115]
 [-0.77503367]
 [ 1.96145159]]
```

4 predict f0r 2025

[ ]:
```
import pandas as pd
import joblib
from sklearn.preprocessing import StandardScaler

# Load the trained model and scaler
rf = joblib.load('rf_model.pkl')  # Load the Random Forest model
standardScaler = joblib.load('scaler.pkl')  # Load the StandardScaler if used

# Load the dataset for 2024 and 2025 (assuming you have datasets containing␣
 ↪2024 and 2025 data)
dataset_2024 = pd.read_csv("dataset.csv")  # Replace with actual path to the␣
 ↪2024 dataset
dataset_2025 = pd.read_csv("car predict 2025 assum.csv")  # Corrected filename

# Print column names to inspect the dataset
print("Columns in dataset_2024:", dataset_2024.columns)
print("Columns in dataset_2025:", dataset_2025.columns)

# Drop non-numeric columns like 'Brand' and 'Model' if they exist in both␣
 ↪datasets
columns_to_drop = ['Brand', 'Model']
for col in columns_to_drop:
    if col in dataset_2024.columns:
        dataset_2024 = dataset_2024.drop(col, axis=1)  # Drop column and␣
 ↪reassign to avoid chained assignment
    if col in dataset_2025.columns:
        dataset_2025 = dataset_2025.drop(col, axis=1)  # Drop column and␣
 ↪reassign to avoid chained assignment

# Clean the 'Fuel_Efficiency' column (or any other similar column)
for col in ['Fuel_Efficiency']:  # Replace with the actual column name causing␣
 ↪the issue
```

```
    if col in dataset_2024.columns:
        # Convert to string and clean, handling any non-numeric characters
        dataset_2024[col] = dataset_2024[col].astype(str).str.replace(r'[^0-9.
 ↪]', '', regex=True)
        dataset_2024[col] = pd.to_numeric(dataset_2024[col], errors='coerce')
        dataset_2024[col] = dataset_2024[col].fillna(dataset_2024[col].mean()) ␣
 ↪# Fill NaNs with the mean value

    if col in dataset_2025.columns:
        dataset_2025[col] = dataset_2025[col].astype(str).str.replace(r'[^0-9.
 ↪]', '', regex=True)
        dataset_2025[col] = pd.to_numeric(dataset_2025[col], errors='coerce')
        dataset_2025[col] = dataset_2025[col].fillna(dataset_2025[col].mean()) ␣
 ↪# Fill NaNs with the mean value

# Clean the 'Mileage' column in both datasets
for col in ['Mileage']:
    if col in dataset_2024.columns:
        dataset_2024[col] = dataset_2024[col].astype(str).str.replace(r'[^0-9.
 ↪]', '', regex=True)
        dataset_2024[col] = pd.to_numeric(dataset_2024[col], errors='coerce')
        dataset_2024[col] = dataset_2024[col].fillna(dataset_2024[col].mean()) ␣
 ↪# Fill NaNs with the mean value
    if col in dataset_2025.columns:
        dataset_2025[col] = dataset_2025[col].astype(str).str.replace(r'[^0-9.
 ↪]', '', regex=True)
        dataset_2025[col] = pd.to_numeric(dataset_2025[col], errors='coerce')
        dataset_2025[col] = dataset_2025[col].fillna(dataset_2025[col].mean()) ␣
 ↪# Fill NaNs with the mean value

# Adjust the 'Year' column for prediction
dataset_2024['Year'] = dataset_2024['Year'].apply(lambda x: 2025 - x)  #␣
 ↪Adjusting the year for prediction
dataset_2025['Year'] = dataset_2025['Year'].apply(lambda x: 2025 - x)  #␣
 ↪Adjusting the year for prediction

# Process 'Fuel_Type' column (split if needed)
for dataset in [dataset_2024, dataset_2025]:
    make_train = dataset["Fuel_Type"].str.split(" ", expand=True)
    dataset["Fuel_Type"] = make_train[0]

# Apply one-hot encoding for 'Fuel_Type' and 'Transmission' for both 2024 and␣
 ↪2025
dataset_2024 = pd.get_dummies(dataset_2024, columns=["Fuel_Type",␣
 ↪"Transmission"], drop_first=True)
```

```
dataset_2025 = pd.get_dummies(dataset_2025, columns=["Fuel_Type",␣
 ↪"Transmission"], drop_first=True)

# Ensure the columns are in the same order for both datasets
model_features = joblib.load('model_features.pkl')  # Load the saved feature␣
 ↪names

# Align the datasets with the same feature columns
dataset_2024 = dataset_2024[model_features]
dataset_2025 = dataset_2025[model_features]

# Define features (X) for 2024 and 2025 data (without the target column 'Price')
X_2024 = dataset_2024
X_2025 = dataset_2025

# Scale the features using the pre-fitted scaler
X_2024_scaled = standardScaler.transform(X_2024)
X_2025_scaled = standardScaler.transform(X_2025)

# Ensure that X_2024 and X_2025 have feature names
X_2024_scaled = pd.DataFrame(X_2024_scaled, columns=X_2024.columns)
X_2025_scaled = pd.DataFrame(X_2025_scaled, columns=X_2025.columns)

# Predict the prices for the 2024 and 2025 data
predicted_prices_2024 = rf.predict(X_2024_scaled)
predicted_prices_2025 = rf.predict(X_2025_scaled)

# Compare predicted prices for both years
comparison_2024 = pd.DataFrame({
    'Year': [2024] * len(predicted_prices_2024),
    'Predicted Price': predicted_prices_2024
})

comparison_2025 = pd.DataFrame({
    'Year': [2025] * len(predicted_prices_2025),
    'Predicted Price': predicted_prices_2025
})

# Combine both years into one DataFrame for comparison
comparison = pd.concat([comparison_2024, comparison_2025])

# Save the comparison results to a CSV file
comparison.to_csv('price_comparison_2024_2025.csv', index=False)

# Display the comparison for both years
print("Price Comparison for 2024 and 2025:")
print(comparison.head())  # Display a few rows of the comparison
```

```
Columns in dataset_2024: Index(['Unnamed: 0', 'Name', 'Location', 'Year',
'Kilometers_Driven',
       'Fuel_Type', 'Transmission', 'Owner_Type', 'Mileage', 'Engine', 'Power',
       'Seats', 'New_Price', 'Price'],
      dtype='object')
Columns in dataset_2025: Index(['Brand', 'Model', 'Year', 'Mileage',
'Fuel_Type', 'Transmission',
       'Price'],
      dtype='object')
Price Comparison for 2024 and 2025:
   Year  Predicted Price
0  2024          42560.0
1  2024          42560.0
2  2024          36070.0
3  2024          42560.0
4  2024          42640.0
```

5 comparsion in price difference of 2024 and 2025 year

[ ]:
```
# Merge the two DataFrames using their index
merged_comparison = pd.merge(comparison_2024, comparison_2025, left_index=True,␣
 ↪right_index=True, suffixes=('_2024', '_2025'))

# Calculate the difference in prices between 2025 and 2024
merged_comparison['Price Difference'] = merged_comparison['Predicted␣
 ↪Price_2025'] - merged_comparison['Predicted Price_2024']

# Calculate the percentage rise in 2025 car price
merged_comparison['Percentage Rise'] = ((merged_comparison['Predicted␣
 ↪Price_2025'] - merged_comparison['Predicted Price_2024']) /␣
 ↪merged_comparison['Predicted Price_2024']) * 100

# Display the DataFrame with Percentage Rise
print(merged_comparison[['Predicted Price_2024', 'Predicted Price_2025', 'Price␣
 ↪Difference', 'Percentage Rise']])
comparison.to_csv('difference in 2024 and 2025.csv', index=False)
```

```
   Predicted Price_2024  Predicted Price_2025  Price Difference  \
0               42560.0               36730.0           -5830.0
1               42560.0               42560.0               0.0
2               36070.0               36070.0               0.0
3               42560.0               42640.0              80.0
4               42640.0               36730.0           -5910.0
5               42560.0               42640.0              80.0
6               42560.0               36070.0           -6490.0
7               42640.0               36730.0           -5910.0
8               42560.0               42560.0               0.0
```

```
9              42560.0              36730.0           -5830.0
```

```
   Percentage Rise
0       -13.698308
1         0.000000
2         0.000000
3         0.187970
4       -13.860225
5         0.187970
6       -15.249060
7       -13.860225
8         0.000000
9       -13.698308
```

[ ]: