**Tasks**

**[Main Task]**

**1. Build a model to predict car prices (maximize profit for Autochek and comfortable price for buyers)**

**[Main Task break down]**

**1. Using pandas, open all projcet related csv (Done)**

**(SQL Queries)**

**[a] (create db) --> CREATE DATABASE autochekdb**

WITH

OWNER = postgres

ENCODING = 'UTF8'

LC\_COLLATE = 'en\_NG'

LC\_CTYPE = 'en\_NG'

TABLESPACE = pg\_default

CONNECTION LIMIT = -1;

**[b] (create table bodytype) --> create table bodytype (**

id int,

title varchar(255),

description varchar(255),

sailthru\_tag varchar(255)

**);**

**[c] (copy data from csv to bodytype table) -->** \copy bodytype

FROM '/home/vcode/Documents/data science project/AutoChek Africa Interview Assessment/data/docs copy/bodytype.csv'

DELIMITER ';'

CSV HEADER;

**[d] (create table categories) --> create table categories (**

id int,

title varchar(255),

slug varchar(255),

price\_guide int

**);**

**[e] (copy data from csv to categories table) -->** \copy categories

FROM '/home/vcode/Documents/data science project/AutoChek Africa Interview Assessment/data/docs copy/categories.csv'

DELIMITER ';'

CSV HEADER;

**[f] (create table condition) --> create table condition (**

id int,

title varchar(255),

description varchar(255),

**);**

**[g] (copy data from csv to condition table) -->** \copy condition

FROM '/home/vcode/Documents/data science project/AutoChek Africa Interview Assessment/data/docs copy/condition.csv'

DELIMITER ';'

CSV HEADER;

**[h] (create table listing) --> create table listing (**

id int,

old\_id int,

title varchar(255),

location\_id int,

listing\_status\_id int

**);**

**[i] (copy data from csv to listing table) -->** \copy listing

FROM '/home/vcode/Documents/data science project/AutoChek Africa Interview Assessment/data/docs copy/listing.csv'

DELIMITER ';'

CSV HEADER;

**[j] (create table trueprices) --> create table trueprices (**

id int,

make\_id int,

model\_id int,

series\_id int,

is\_verified\_dealer int,

price varchar(255),

year\_of\_manufacture int,

domain\_id int,

listing\_id int,

condition\_type\_id int,

body\_type\_id int

**);**

**[k] (copy data from csv to trueprices table) -->** \copy trueprices

FROM '/home/vcode/Documents/data science project/AutoChek Africa Interview Assessment/data/docs copy/trueprices.csv'

DELIMITER ';'

CSV HEADER;

**2. Join by ID (Also use sql to join the data by their ID) (Done)**

**(SQL Queries)**

**select**

**tp.id,**

**tp.make\_id,**

**tp.model\_id,**

**tp.series\_id,**

**tp.is\_verified\_dealer,**

**tp.price,**

**tp.year\_of\_manufacture,**

**l.title as listingTitle,**

**ct.title as conditionTitle,**

**bt.sailthru\_tag**

**from trueprices tp**

**left join listing l**

**on tp.listing\_id = l.id**

**left join bodytype bt**

**on tp.body\_type\_id = bt.id**

**left join condition ct**

**on tp.condition\_type\_id = ct.id**

**order by id**

**3. Clean data after join is complete (Done)**

**4. predict car prices (Done)**

**5. Evaluate performance using RMSE (Done)**

**6. Build a Dashboard studio using** **DataStudio with the clean data to present the data analysis. (Done)**

**7. Write a brief report on steps taken to solve the prediction problem, findings from data analysis and result of the model. (Done)**