ST1516 – DevOps & AI Automation
Jeyakumar Sriram
DAAA/FT/2B/01
2214618

# Giggles Gears

Where you can find you future car...

# Table of contents

01

Application

02

DevOps Process

03

Regression

04

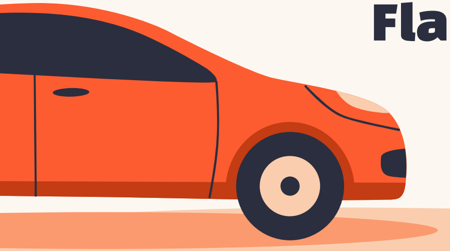
Flask

05

Testing

06

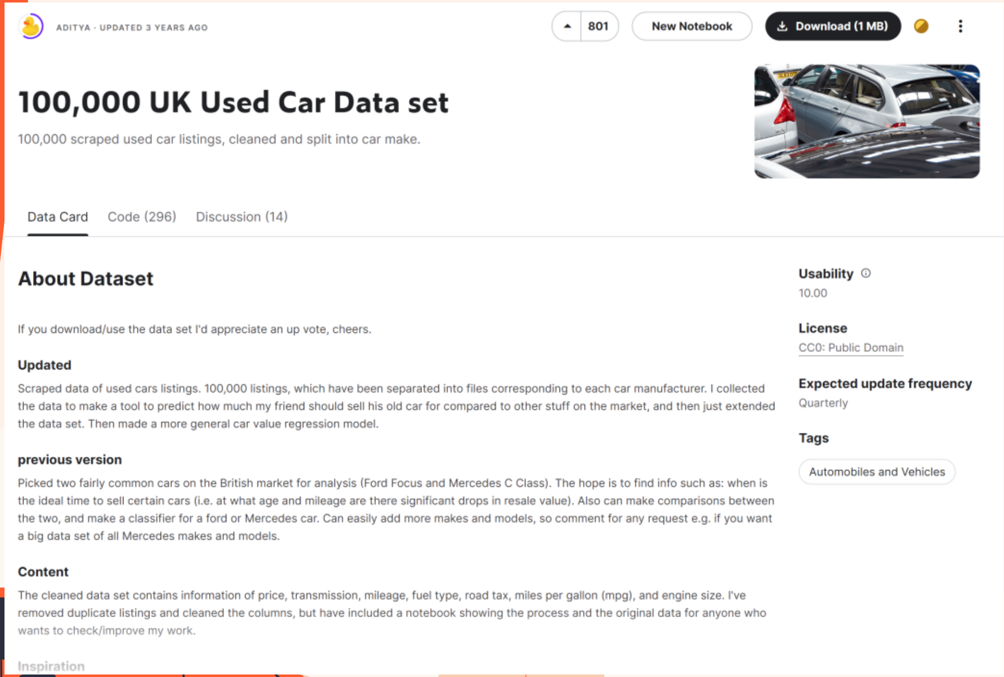
Advanced



# 01

# Application

# The Dataset



## 100,000 UK Used Car Data set

100,000 scraped used car listings, cleaned and split into car make.

### About Dataset

If you download/use the data set I'd appreciate an up vote, cheers.

**Updated**

Scraped data of used cars listings. 100,000 listings, which have been separated into files corresponding to each car manufacturer. I collected the data to make a tool to predict how much my friend should sell his old car for compared to other stuff on the market, and then just extended the data set. Then made a more general car value regression model.

**previous version**

Picked two fairly common cars on the British market for analysis (Ford Focus and Mercedes C Class). The hope is to find info such as: when is the ideal time to sell certain cars (i.e. at what age and mileage are there significant drops in resale value). Also can make comparisons between the two, and make a classifier for a ford or Mercedes car. Can easily add more makes and models, so comment for any request e.g. if you want a big data set of all Mercedes makes and models.

**Content**

The cleaned data set contains information of price, transmission, mileage, fuel type, road tax, miles per gallon (mpg), and engine size. I've removed duplicate listings and cleaned the columns, but have included a notebook showing the process and the original data for anyone who wants to check/improve my work.

Inspiration

**Usability**

10.00

**License**

CC0: Public Domain

**Expected update frequency**

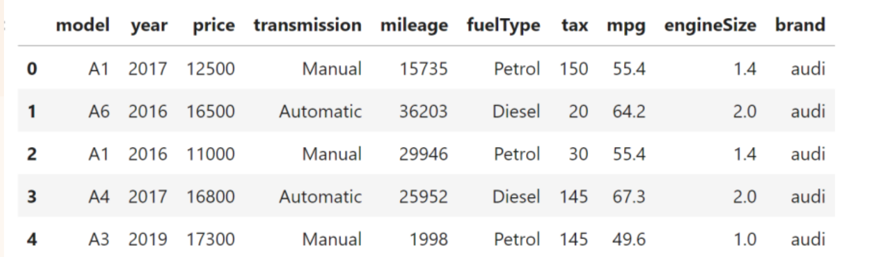
Quarterly

**Tags**

Automobiles and Vehicles

This dataset consists of prices and other related information on a 100000 cars from the UK. Since this dataset is from the UK most of the brands are from there.

I had to join and clean them to properly format to one df



| | model | year | price | transmission | mileage | fuelType | tax | mpg | engineSize | brand |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | A1 | 2017 | 12500 | Manual | 15735 | Petrol | 150 | 55.4 | 1.4 | audi |
| **1** | A6 | 2016 | 16500 | Automatic | 36203 | Diesel | 20 | 64.2 | 2.0 | audi |
| **2** | A1 | 2016 | 11000 | Manual | 29946 | Petrol | 30 | 55.4 | 1.4 | audi |
| **3** | A4 | 2017 | 16800 | Automatic | 25952 | Diesel | 145 | 67.3 | 2.0 | audi |
| **4** | A3 | 2019 | 17300 | Manual | 1998 | Petrol | 145 | 49.6 | 1.0 | audi |

# Application Demo

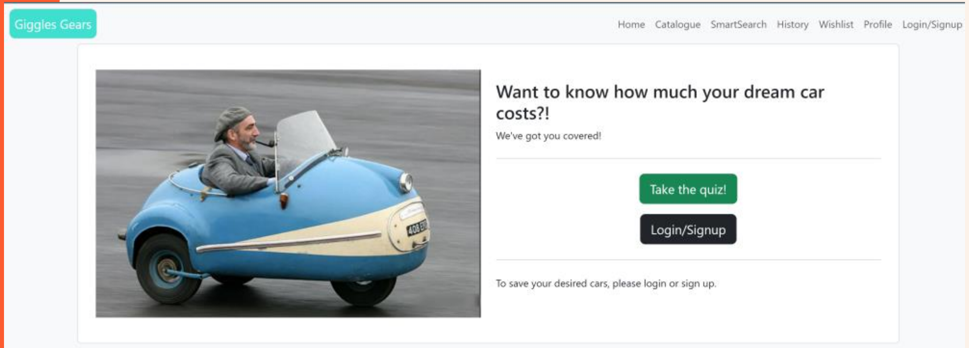
The application I built has the following key features.

Users can:

- browse a catalogue of cars available for resale
- add their desired cars to a wishlist
- get a rough estimate of how much a car with their desired features would cost
- get 3 recommended cars based on their input on desired features
- view and delete their prediction history
- view and delete items from wishlist
- sign up or login to save their progress
- view and modify their profile

https://giggles-gears.onrender.com/

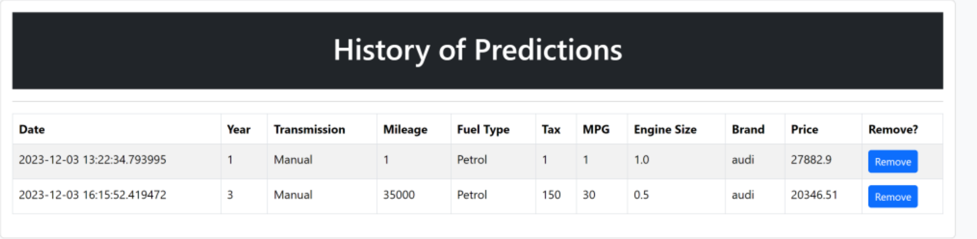
# Some pictures



Want to know how much your dream car costs?!

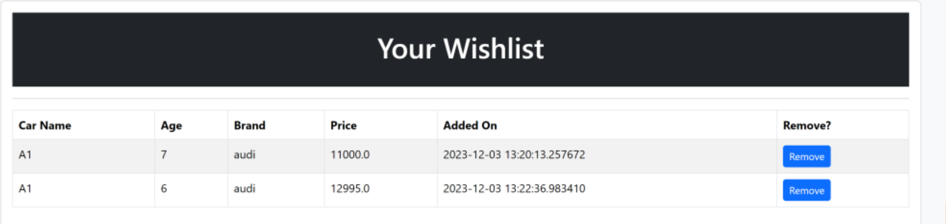
We've got you covered!

To save your desired cars, please login or sign up.



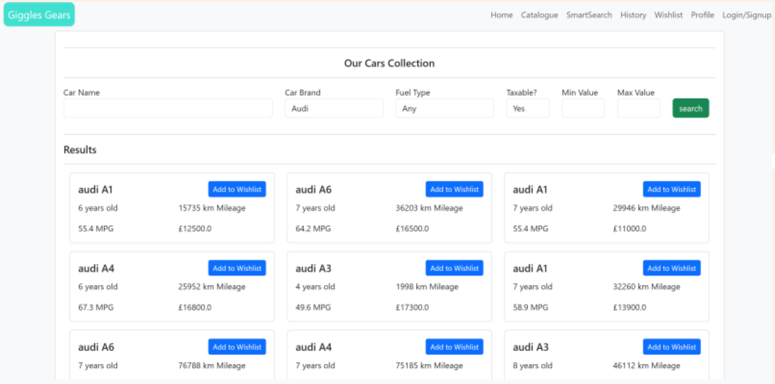
## History of Predictions

| Date | Year | Transmission | Mileage | Fuel Type | Tax | MPG | Engine Size | Brand | Price | Remove? |
|---|---|---|---|---|---|---|---|---|---|---|
| 2023-12-03 13:22:34.793995 | 1 | Manual | 1 | Petrol | 1 | 1 | 1.0 | audi | 27882.9 | Remove |
| 2023-12-03 16:15:52.419472 | 3 | Manual | 35000 | Petrol | 150 | 30 | 0.5 | audi | 20346.51 | Remove |



## Your Wishlist

| Car Name | Age | Brand | Price | Added On | Remove? |
|---|---|---|---|---|---|
| A1 | 7 | audi | 11000.0 | 2023-12-03 13:20:13.257672 | Remove |
| A1 | 6 | audi | 12995.0 | 2023-12-03 13:22:36.983410 | Remove |



Our Cars Collection

Car Name:
Car Brand: Audi
Fuel Type: Any
Taxable?: Yes
Min Value:
Max Value:

Results

audi A1 — 6 years old — 15735 km Mileage — 55.4 MPG — £12500.0

audi A6 — 7 years old — 36203 km Mileage — 64.2 MPG — £16500.0

audi A1 — 7 years old — 29946 km Mileage — 55.4 MPG — £11000.0

audi A4 — 6 years old — 25952 km Mileage — 67.3 MPG — £16800.0

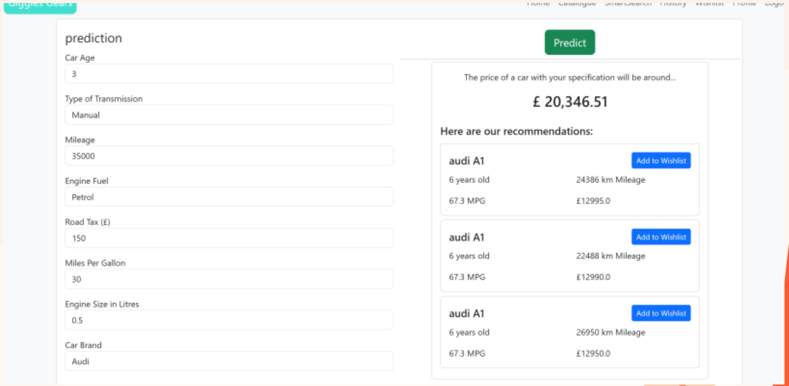
audi A3 — 4 years old — 1998 km Mileage — 49.6 MPG — £17300.0

audi A1 — 7 years old — 32260 km Mileage — 58.9 MPG — £13900.0

audi A6 — 7 years old — 76788 km Mileage

audi A4 — 7 years old — 75185 km Mileage

audi A3 — 8 years old — 46112 km Mileage



prediction

Car Age: 3

Type of Transmission: Manual

Mileage: 35000

Engine Fuel: Petrol

Road Tax (£): 150

Miles Per Gallon: 30

Engine Size in Litres: 0.5

Car Brand: Audi

The price of a car with your specification will be around...

£ 20,346.51

Here are our recommendations:

audi A1 — 6 years old — 24386 km Mileage — 67.3 MPG — £12995.0

audi A1 — 6 years old — 22488 km Mileage — 67.3 MPG — £12990.0

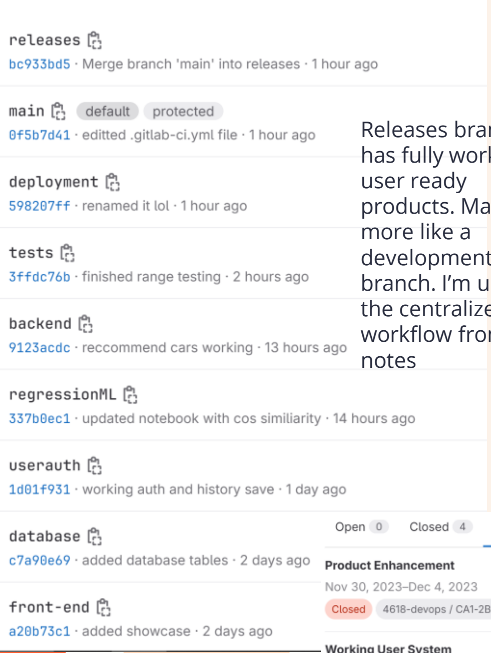
audi A1 — 6 years old — 26950 km Mileage — 67.3 MPG — £12950.0

## 02

# DevOps Process

# Git & Scrum

## Branches



releases
bc933bd5 · Merge branch 'main' into releases · 1 hour ago

main
0f5b7d41 · editted .gitlab-ci.yml file · 1 hour ago

deployment
598207ff · renamed it lol · 1 hour ago

tests
3ffdc76b · finished range testing · 2 hours ago

backend
9123acdc · reccommend cars working · 13 hours ago

regressionML
337b0ec1 · updated notebook with cos similiarity · 14 hours ago

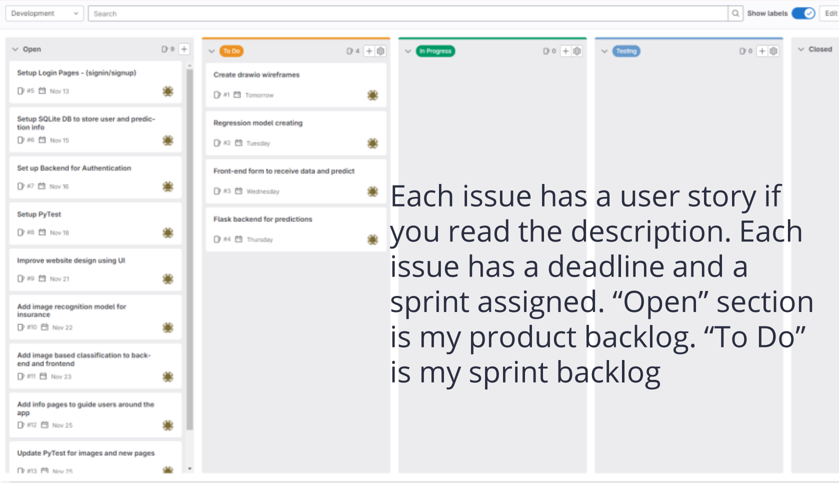
userauth
1d01f931 · working auth and history save · 1 day ago

database
c7a90e69 · added database tables · 2 days ago

front-end
a20b73c1 · added showcase · 2 days ago

Releases branch has fully working user ready products. Main is more like a development branch. I'm using the centralized workflow from notes

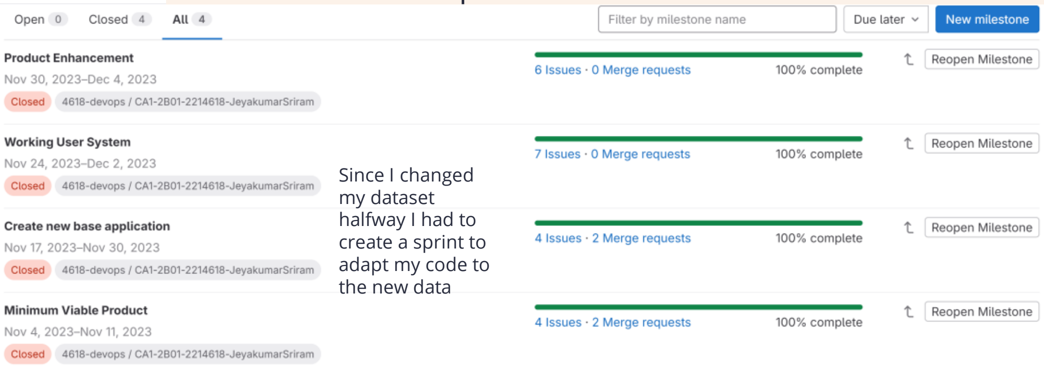
## Issue List -> Scrum Board



| Open 9 | To Do 4 | In Progress 0 | Testing 0 | Closed |
|---|---|---|---|---|
| Setup Login Pages - (signin/signup)<br>#5 Nov 13 | Create drawio wireframes<br>#1 Tomorrow | | | |
| Setup SQLite DB to store user and prediction info<br>#6 Nov 15 | Regression model creating<br>#2 Tuesday | | | |
| Set up Backend for Authentication<br>#7 Nov 16 | Front-end form to receive data and predict<br>#3 Wednesday | | | |
| Setup PyTest<br>#8 Nov 18 | Flask backend for predictions<br>#4 Thursday | | | |
| Improve website design using UI<br>#9 Nov 21 | | | | |
| Add image recognition model for insurance<br>#10 Nov 22 | | | | |
| Add image based classification to backend and frontend<br>#11 Nov 23 | | | | |
| Add info pages to guide users around the app<br>#12 Nov 25 | | | | |
| Update PyTest for images and new pages<br>#13 Nov 25 | | | | |

Each issue has a user story if you read the description. Each issue has a deadline and a sprint assigned. "Open" section is my product backlog. "To Do" is my sprint backlog

## Milestones -> Sprints



**Product Enhancement**
Nov 30, 2023–Dec 4, 2023
Closed · 4618-devops / CA1-2B01-2214618-JeyakumarSriram
6 Issues · 0 Merge requests · 100% complete
Reopen Milestone

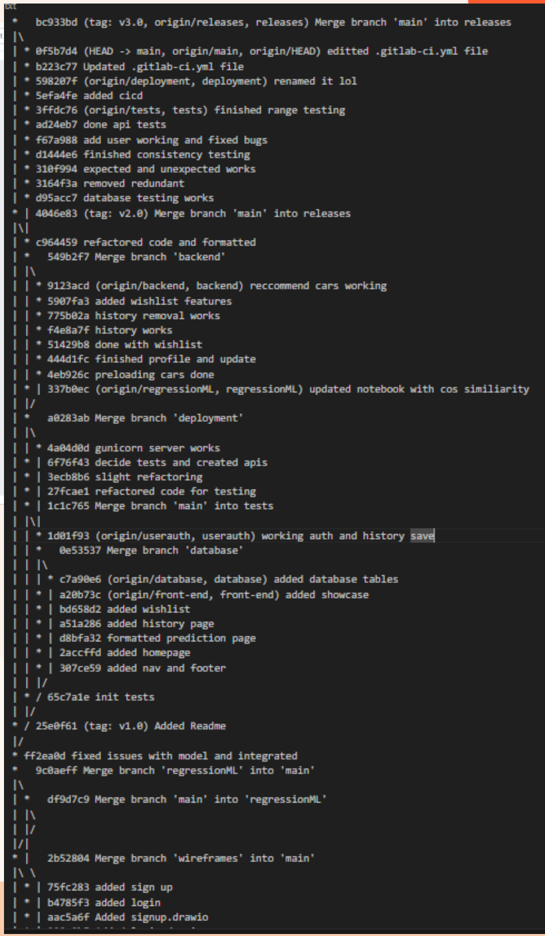
**Working User System**
Nov 24, 2023–Dec 2, 2023
Closed · 4618-devops / CA1-2B01-2214618-JeyakumarSriram
7 Issues · 0 Merge requests · 100% complete
Reopen Milestone

**Create new base application**
Nov 17, 2023–Nov 30, 2023
Closed · 4618-devops / CA1-2B01-2214618-JeyakumarSriram
4 Issues · 2 Merge requests · 100% complete
Reopen Milestone

**Minimum Viable Product**
Nov 4, 2023–Nov 11, 2023
Closed · 4618-devops / CA1-2B01-2214618-JeyakumarSriram
4 Issues · 2 Merge requests · 100% complete
Reopen Milestone

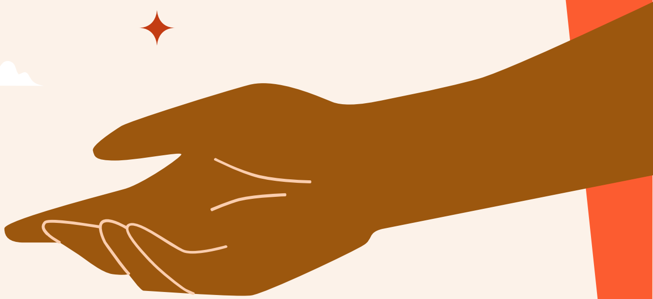
Since I changed my dataset halfway I had to create a sprint to adapt my code to the new data

## Git Commits



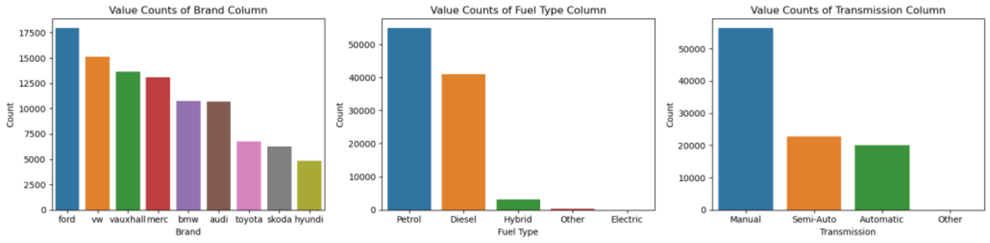
```
*   bc933bd (tag: v3.0, origin/releases, releases) Merge branch 'main' into releases
|\
| * 0f5b7d4 (HEAD -> main, origin/main, origin/HEAD) editted .gitlab-ci.yml file
| * b223c77 Updated .gitlab-ci.yml file
| * 598207f (origin/deployment, deployment) renamed it lol
| * 5efa4fe added cicd
| * 3ffdc76 (origin/tests, tests) finished range testing
| * ad24eb7 done api tests
| * f67a988 add user working and fixed bugs
| * d1444e6 finished consistency testing
| * 310f994 expected and unexpected works
| * 3164f3a removed redundant
| * d95acc7 database testing works
* | 4046e83 (tag: v2.0) Merge branch 'main' into releases
|\|
| * c964459 refactored code and formatted
| *   549b2f7 Merge branch 'backend'
| |\
| | * 9123acd (origin/backend, backend) reccommend cars working
| | * 5907fa3 added wishlist features
| | * 775b02a history removal works
| | * f4e8a7f history works
| | * 51429b8 done with wishlist
| | * 444d1fc finished profile and update
| | * 4eb926c preloading cars done
| * | 337b0ec (origin/regressionML, regressionML) updated notebook with cos similiarity
| |/
| *   a0283ab Merge branch 'deployment'
| |\
| | * 4a04d0d gunicorn server works
| * | 6f76f43 decide tests and created apis
| * | 3ecb8b6 slight refactoring
| * | 27fcae1 refactored code for testing
| * | 1c1c765 Merge branch 'main' into tests
| |\|
| | * 1d01f93 (origin/userauth, userauth) working auth and history save
| | *   0e53537 Merge branch 'database'
| | |\
| | | * c7a90e6 (origin/database, database) added database tables
| | * | a20b73c (origin/front-end, front-end) added showcase
| | * | bd658d2 added wishlist
| | * | a51a286 added history page
| | * | d8bfa32 formatted prediction page
| | * | 2accffd added homepage
| | * | 307ce59 added nav and footer
| | |/
| * / 65c7a1e init tests
| |/
* / 25e0f61 (tag: v1.0) Added Readme
|/
* ff2ea0d fixed issues with model and integrated
*   9c0aeff Merge branch 'regressionML' into 'main'
|\
| *   df9d7c9 Merge branch 'main' into 'regressionML'
| |\
| |/
|/|
* |   2b52804 Merge branch 'wireframes' into 'main'
|\ \
| * | 75fc283 added sign up
| * | b4785f3 added login
| * | aac5a6f Added signup.drawio
```

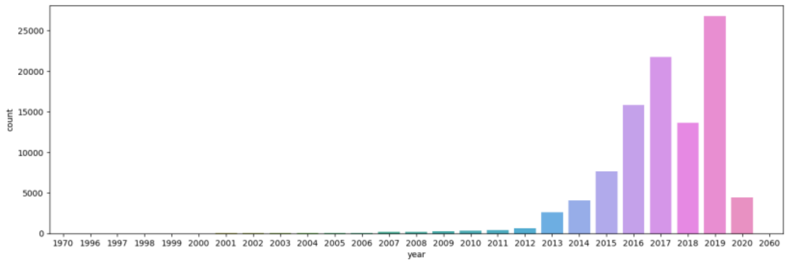
# 03

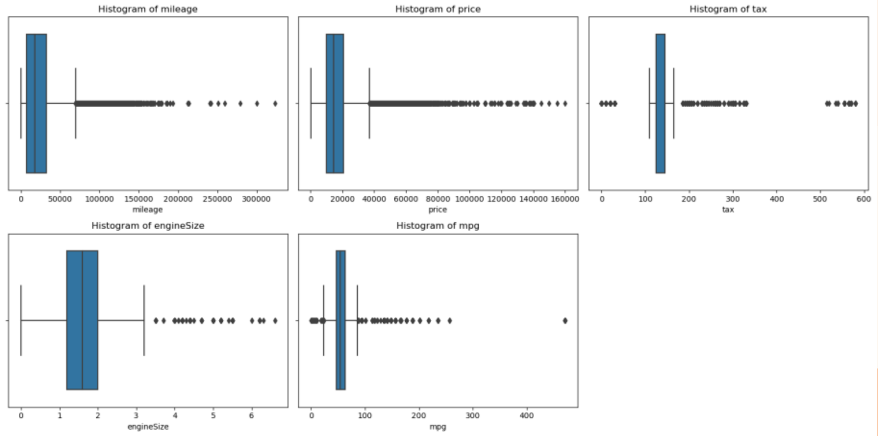


# Regression

# Exploratory Data Analysis

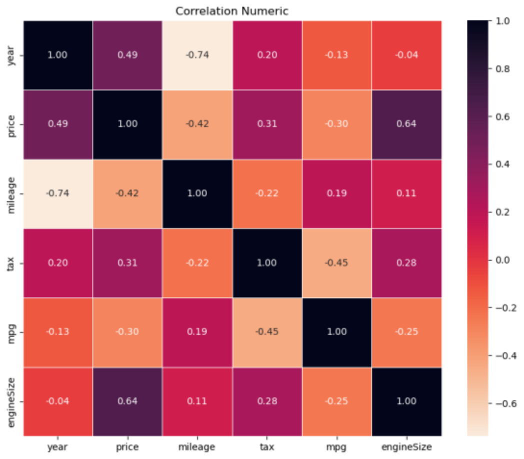






Based on the EDA:

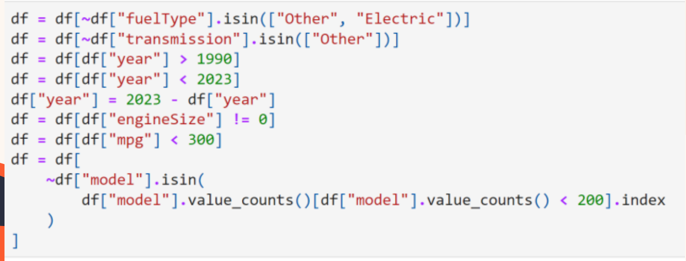
- Some classes had very low rows and decided to remove them.
- Outliers in year and other numeric columns
- Price highly correlated with year and engine size
- No null values
- Model has too many classes and cannot be used



# Preprocessing

## Cleaning

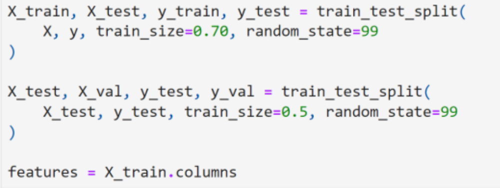
- Removed outliers
- Remove classes with low representation
- Removed models with very little rows



```
df = df[~df["fuelType"].isin(["Other", "Electric"])]
df = df[~df["transmission"].isin(["Other"])]
df = df[df["year"] > 1990]
df = df[df["year"] < 2023]
df["year"] = 2023 - df["year"]
df = df[df["engineSize"] != 0]
df = df[df["mpg"] < 300]
df = df[
    ~df["model"].isin(
        df["model"].value_counts()[df["model"].value_counts() < 200].index
    )
]
```

## Splitting

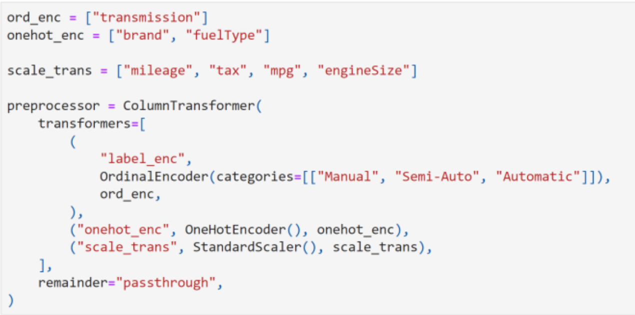
- Used train-val-test split with ratio 70-10-10
- Dropped model column from dataset



```
X_train, X_test, y_train, y_test = train_test_split(
    X, y, train_size=0.70, random_state=99
)

X_test, X_val, y_test, y_val = train_test_split(
    X_test, y_test, train_size=0.5, random_state=99
)

features = X_train.columns
```

## Pipeline

- Performed ordinal encoding for transmission
- Performed one hot for others
- Scaled numeric columns

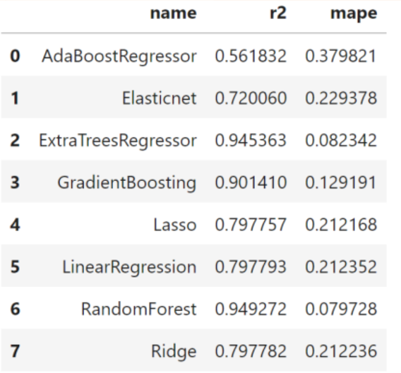


```
ord_enc = ["transmission"]
onehot_enc = ["brand", "fuelType"]

scale_trans = ["mileage", "tax", "mpg", "engineSize"]

preprocessor = ColumnTransformer(
    transformers=[
        (
            "label_enc",
            OrdinalEncoder(categories=[["Manual", "Semi-Auto", "Automatic"]]),
            ord_enc,
        ),
        ("onehot_enc", OneHotEncoder(), onehot_enc),
        ("scale_trans", StandardScaler(), scale_trans),
    ],
    remainder="passthrough",
)
```

# Modelling

## Try many

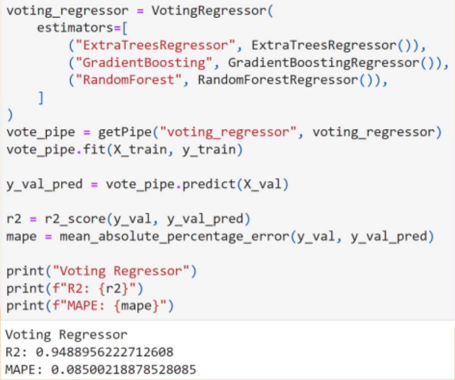
- Compared r2 and MAPE from various models
- Random Forest, ExtraTrees and Gradient Boosting performed well on val



| | name | r2 | mape |
|---|---|---|---|
| 0 | AdaBoostRegressor | 0.561832 | 0.379821 |
| 1 | Elasticnet | 0.720060 | 0.229378 |
| 2 | ExtraTreesRegressor | 0.945363 | 0.082342 |
| 3 | GradientBoosting | 0.901410 | 0.129191 |
| 4 | Lasso | 0.797757 | 0.212168 |
| 5 | LinearRegression | 0.797793 | 0.212352 |
| 6 | RandomForest | 0.949272 | 0.079728 |
| 7 | Ridge | 0.797782 | 0.212236 |

## VotingRegressor

- Used three best models just now to make it
- Performed around the same
- But takes very long to train and predict

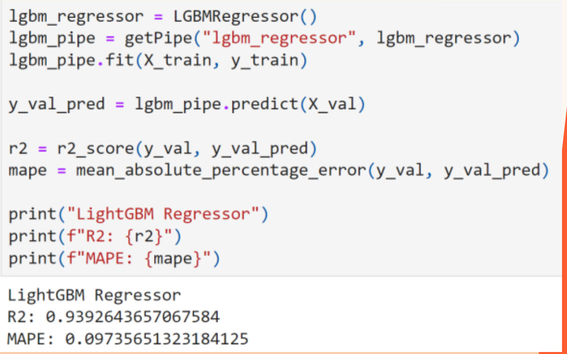


```
voting_regressor = VotingRegressor(
    estimators=[
        ("ExtraTreesRegressor", ExtraTreesRegressor()),
        ("GradientBoosting", GradientBoostingRegressor()),
        ("RandomForest", RandomForestRegressor()),
    ]
)
vote_pipe = getPipe("voting_regressor", voting_regressor)
vote_pipe.fit(X_train, y_train)

y_val_pred = vote_pipe.predict(X_val)

r2 = r2_score(y_val, y_val_pred)
mape = mean_absolute_percentage_error(y_val, y_val_pred)

print("Voting Regressor")
print(f"R2: {r2}")
print(f"MAPE: {mape}")
```

```
Voting Regressor
R2: 0.9488956222712608
MAPE: 0.08500218878528085
```

## LightGBM

- Used from external library
- Performed around the same
- But is noticeably faster



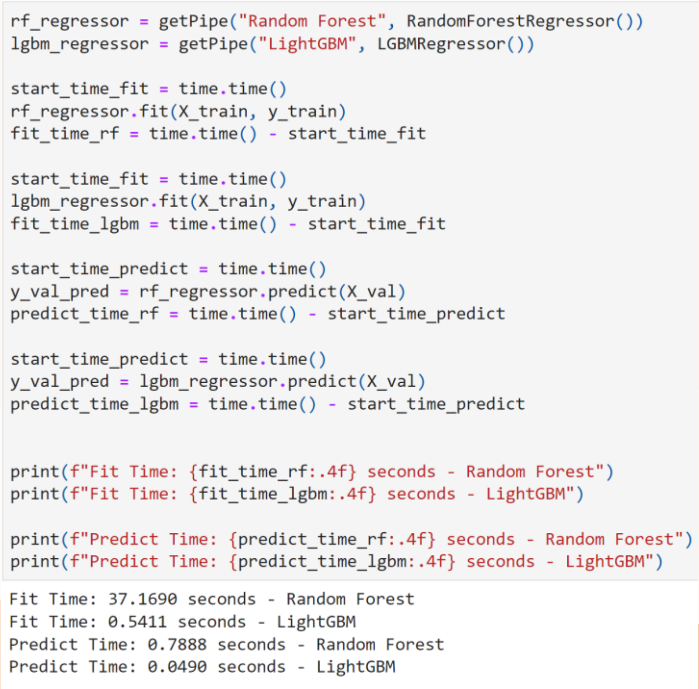
```
lgbm_regressor = LGBMRegressor()
lgbm_pipe = getPipe("lgbm_regressor", lgbm_regressor)
lgbm_pipe.fit(X_train, y_train)

y_val_pred = lgbm_pipe.predict(X_val)

r2 = r2_score(y_val, y_val_pred)
mape = mean_absolute_percentage_error(y_val, y_val_pred)

print("LightGBM Regressor")
print(f"R2: {r2}")
print(f"MAPE: {mape}")
```

```
LightGBM Regressor
R2: 0.9392643657067584
MAPE: 0.09735651323184125
```

# Random Forest vs LightGBM

## Random Forest

- Performed slightly better than LGBM
- Slower than LGBM
- More accuracy might not be that significant to user than speed



```
rf_regressor = getPipe("Random Forest", RandomForestRegressor())
lgbm_regressor = getPipe("LightGBM", LGBMRegressor())

start_time_fit = time.time()
rf_regressor.fit(X_train, y_train)
fit_time_rf = time.time() - start_time_fit

start_time_fit = time.time()
lgbm_regressor.fit(X_train, y_train)
fit_time_lgbm = time.time() - start_time_fit

start_time_predict = time.time()
y_val_pred = rf_regressor.predict(X_val)
predict_time_rf = time.time() - start_time_predict

start_time_predict = time.time()
y_val_pred = lgbm_regressor.predict(X_val)
predict_time_lgbm = time.time() - start_time_predict


print(f"Fit Time: {fit_time_rf:.4f} seconds - Random Forest")
print(f"Fit Time: {fit_time_lgbm:.4f} seconds - LightGBM")

print(f"Predict Time: {predict_time_rf:.4f} seconds - Random Forest")
print(f"Predict Time: {predict_time_lgbm:.4f} seconds - LightGBM")
```

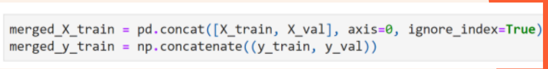
```
Fit Time: 37.1690 seconds - Random Forest
Fit Time: 0.5411 seconds - LightGBM
Predict Time: 0.7888 seconds - Random Forest
Predict Time: 0.0490 seconds - LightGBM
```

## LightGBM

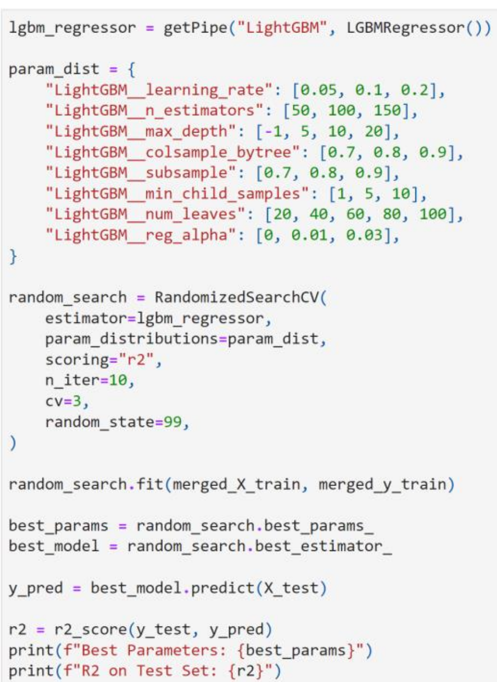
- Performed slightly worse than RF
- 40x faster at training
- 20x faster at prediction
- Will provide better user experience

# Tuning & Final Model

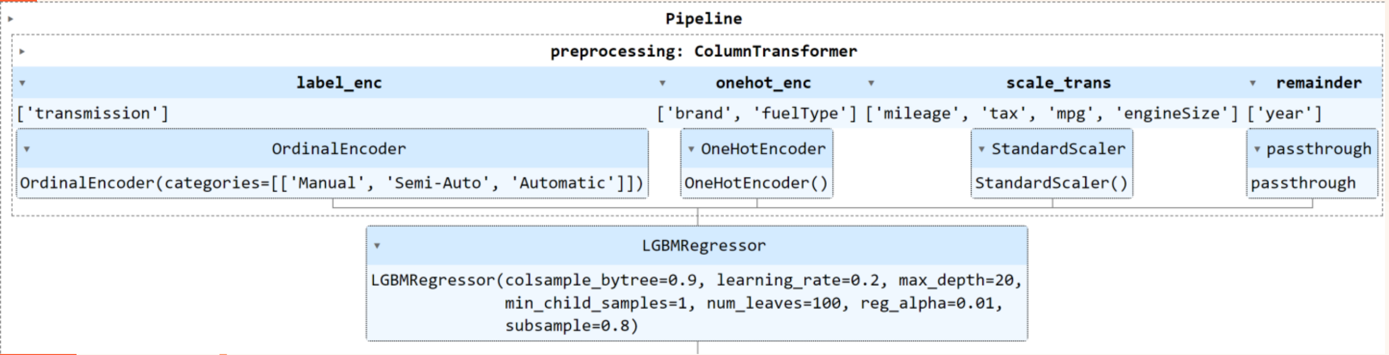
- Since we making final model I will use the entire dataset without splits
- Use RandomisedCV for speed as the slight increase in accuracy with GridsearchCV may not be significant
- Best r2 is better than RF just now
- Final pipeline is ready and will be saved as pickle file to be loaded by Flask



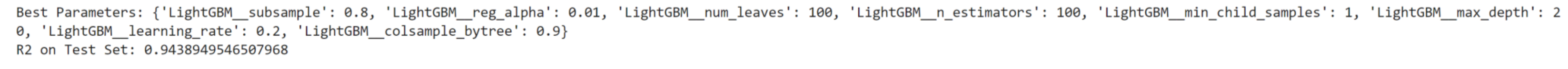
```
merged_X_train = pd.concat([X_train, X_val], axis=0, ignore_index=True)
merged_y_train = np.concatenate((y_train, y_val))
```



```
lgbm_regressor = getPipe("LightGBM", LGBMRegressor())

param_dist = {
    "LightGBM__learning_rate": [0.05, 0.1, 0.2],
    "LightGBM__n_estimators": [50, 100, 150],
    "LightGBM__max_depth": [-1, 5, 10, 20],
    "LightGBM__colsample_bytree": [0.7, 0.8, 0.9],
    "LightGBM__subsample": [0.7, 0.8, 0.9],
    "LightGBM__min_child_samples": [1, 5, 10],
    "LightGBM__num_leaves": [20, 40, 60, 80, 100],
    "LightGBM__reg_alpha": [0, 0.01, 0.03],
}

random_search = RandomizedSearchCV(
    estimator=lgbm_regressor,
    param_distributions=param_dist,
    scoring="r2",
    n_iter=10,
    cv=3,
    random_state=99,
)

random_search.fit(merged_X_train, merged_y_train)

best_params = random_search.best_params_
best_model = random_search.best_estimator_

y_pred = best_model.predict(X_test)

r2 = r2_score(y_test, y_pred)
print(f"Best Parameters: {best_params}")
print(f"R2 on Test Set: {r2}")
```



```
Pipeline
preprocessing: ColumnTransformer
label_enc                    onehot_enc              scale_trans                                remainder
['transmission']             ['brand', 'fuelType']   ['mileage', 'tax', 'mpg', 'engineSize']    ['year']
OrdinalEncoder               OneHotEncoder           StandardScaler                             passthrough
OrdinalEncoder(categories=[['Manual', 'Semi-Auto', 'Automatic']])   OneHotEncoder()   StandardScaler()   passthrough

LGBMRegressor
LGBMRegressor(colsample_bytree=0.9, learning_rate=0.2, max_depth=20,
              min_child_samples=1, num_leaves=100, reg_alpha=0.01,
              subsample=0.8)
```

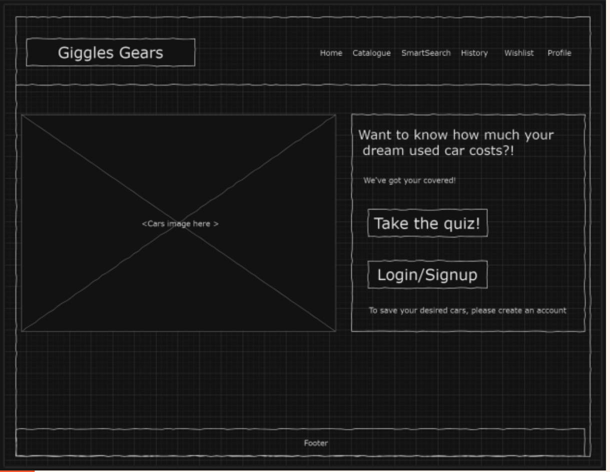


```
Best Parameters: {'LightGBM__subsample': 0.8, 'LightGBM__reg_alpha': 0.01, 'LightGBM__num_leaves': 100, 'LightGBM__n_estimators': 100, 'LightGBM__min_child_samples': 1, 'LightGBM__max_depth': 2
0, 'LightGBM__learning_rate': 0.2, 'LightGBM__colsample_bytree': 0.9}
R2 on Test Set: 0.9438949546507968
```

# 04

# Flask

# Wireframes



Giggles Gears

Home Catalogue SmartSearch History Wishlist Profile

<Cars image here >

Want to know how much your dream used car costs?!

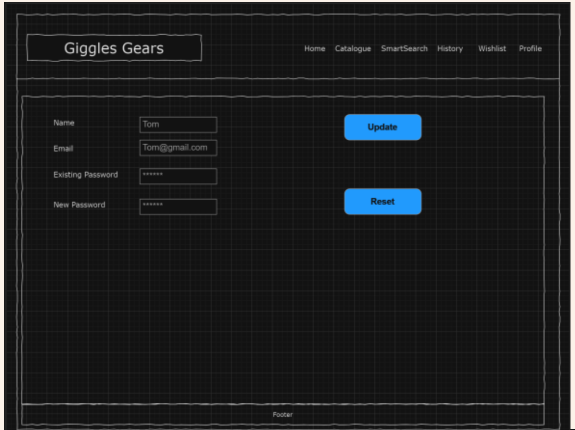
We've got your covered!

Take the quiz!

Login/Signup

To save your desired cars, please create an account

Footer



Giggles Gears

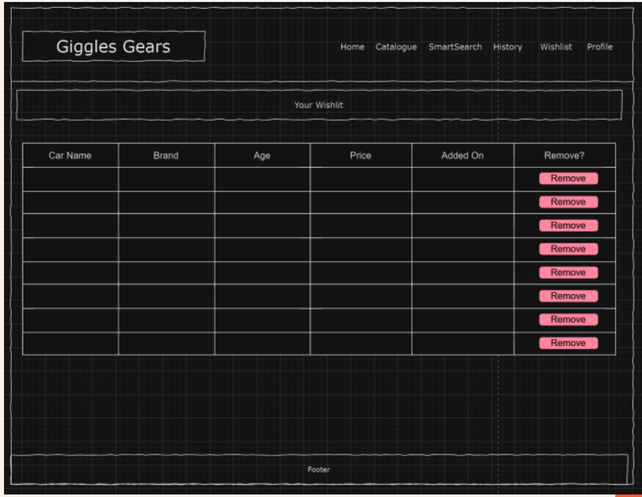
Home Catalogue SmartSearch History Wishlist Profile

|  |  |
|---|---|
| Name | Tom |
| Email | Tom@gmail.com |
| Existing Password | ****** |
| New Password | ****** |

Update

Reset

Footer



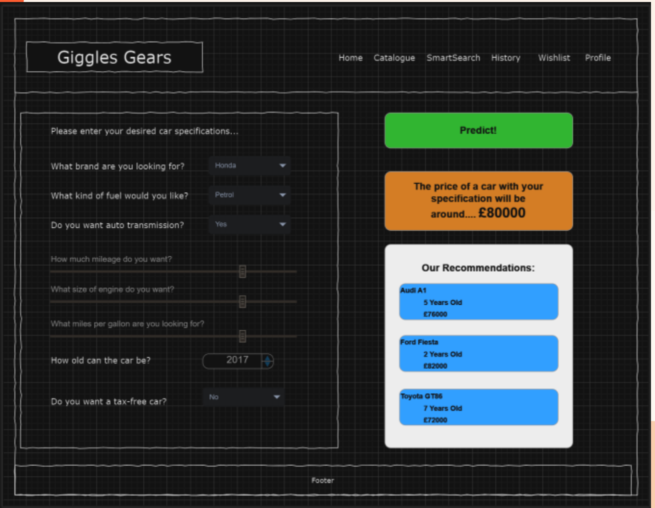
Giggles Gears

Home Catalogue SmartSearch History Wishlist Profile

Your Wishlit

| Car Name | Brand | Age | Price | Added On | Remove? |
|---|---|---|---|---|---|
|  |  |  |  |  | Remove |
|  |  |  |  |  | Remove |
|  |  |  |  |  | Remove |
|  |  |  |  |  | Remove |
|  |  |  |  |  | Remove |
|  |  |  |  |  | Remove |
|  |  |  |  |  | Remove |
|  |  |  |  |  | Remove |

Footer



Giggles Gears

Home Catalogue SmartSearch History Wishlist Profile

Please enter your desired car specifications...

What brand are you looking for? Honda

What kind of fuel would you like? Petrol

Do you want auto transmission? Yes

How much mileage do you want?

What size of engine do you want?

What miles per gallon are you looking for?

How old can the car be? 2017

Do you want a tax-free car? No

Predict!

The price of a car with your specification will be around.... £80000

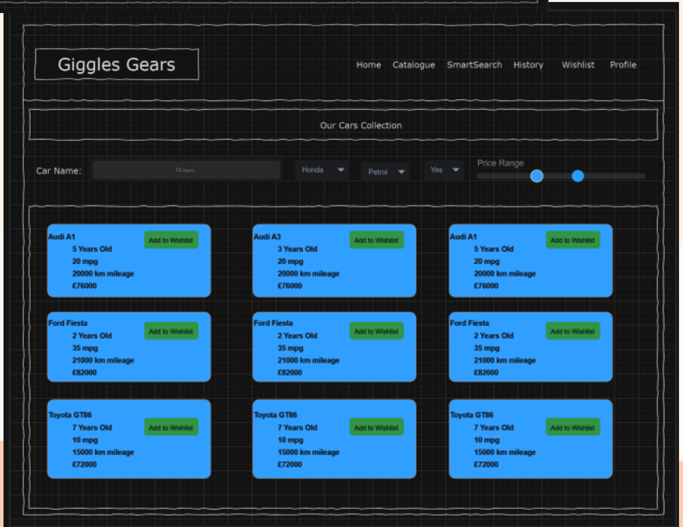
Our Recommendations:

Audi A1
5 Years Old
£76000

Ford Fiesta
2 Years Old
£82000

Toyota GT86
7 Years Old
£72000

Footer



Giggles Gears

Home Catalogue SmartSearch History Wishlist Profile

Our Cars Collection

Car Name: Honda Petrol Yes Price Range

Audi A1
5 Years Old
20 mpg
20000 km mileage
£76000
Add to Wishlist

Audi A3
3 Years Old
20 mpg
20000 km mileage
£76000
Add to Wishlist

Audi A1
5 Years Old
20 mpg
20000 km mileage
£76000
Add to Wishlist

Ford Fiesta
2 Years Old
35 mpg
21000 km mileage
£82000
Add to Wishlist

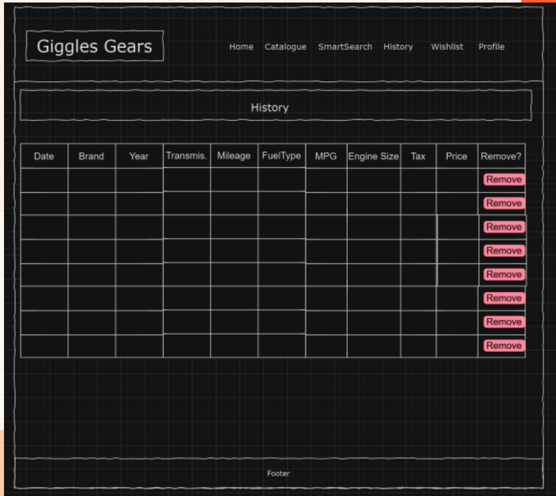
Ford Fiesta
2 Years Old
35 mpg
21000 km mileage
£82000
Add to Wishlist

Ford Fiesta
2 Years Old
35 mpg
21000 km mileage
£82000
Add to Wishlist

Toyota GT86
7 Years Old
10 mpg
15000 km mileage
£72000
Add to Wishlist

Toyota GT86
7 Years Old
10 mpg
15000 km mileage
£72000
Add to Wishlist

Toyota GT86
7 Years Old
10 mpg
15000 km mileage
£72000
Add to Wishlist



Giggles Gears

Home Catalogue SmartSearch History Wishlist Profile

History

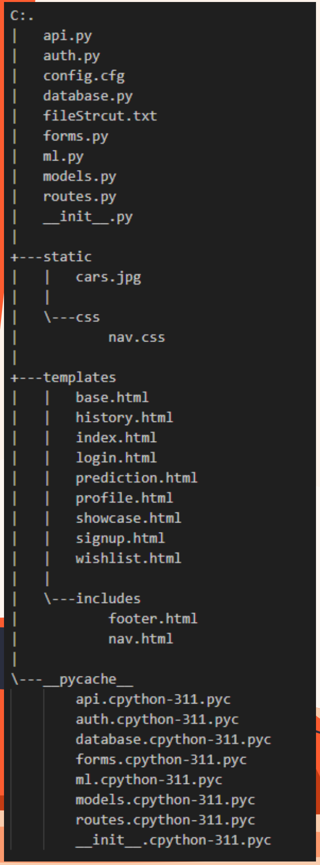
| Date | Brand | Year | Transmis. | Mileage | FuelType | MPG | Engine Size | Tax | Price | Remove? |
|---|---|---|---|---|---|---|---|---|---|---|
|  |  |  |  |  |  |  |  |  |  | Remove |
|  |  |  |  |  |  |  |  |  |  | Remove |
|  |  |  |  |  |  |  |  |  |  | Remove |
|  |  |  |  |  |  |  |  |  |  | Remove |
|  |  |  |  |  |  |  |  |  |  | Remove |
|  |  |  |  |  |  |  |  |  |  | Remove |
|  |  |  |  |  |  |  |  |  |  | Remove |
|  |  |  |  |  |  |  |  |  |  | Remove |

Footer

# Flask App Setup

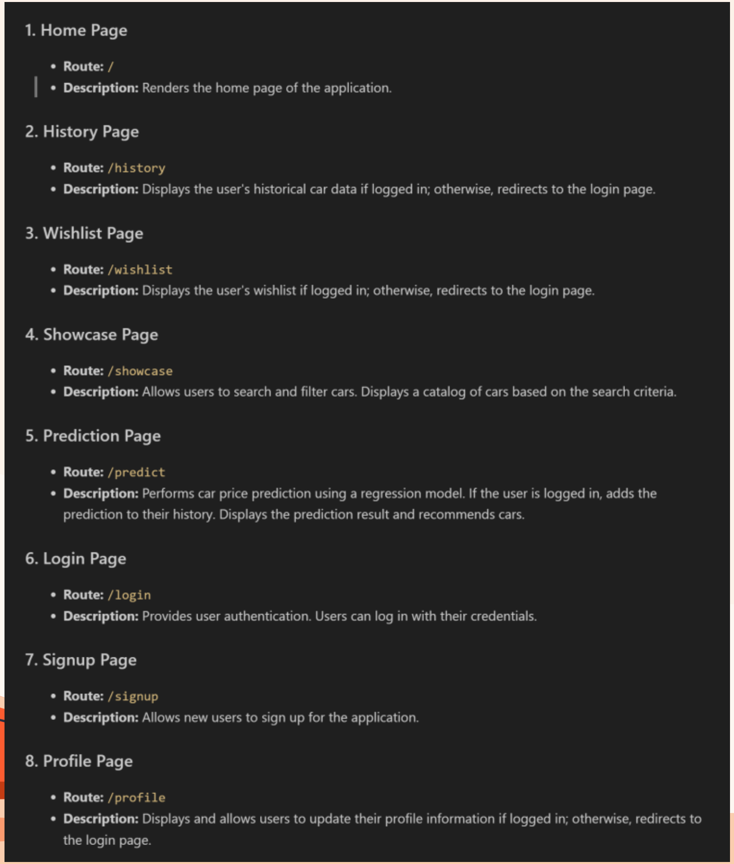
I used Flask-Login for user management and Flask-Bcrypt for hashing passwords

## Application Structure



```
C:.
|   api.py
|   auth.py
|   config.cfg
|   database.py
|   fileStrcut.txt
|   forms.py
|   ml.py
|   models.py
|   routes.py
|   __init__.py
|
+---static
|   |   cars.jpg
|   |
|   \---css
|           nav.css
|
+---templates
|   |   base.html
|   |   history.html
|   |   index.html
|   |   login.html
|   |   prediction.html
|   |   profile.html
|   |   showcase.html
|   |   signup.html
|   |   wishlist.html
|   |
|   \---includes
|           footer.html
|           nav.html
|
\---__pycache__
        api.cpython-311.pyc
        auth.cpython-311.pyc
        database.cpython-311.pyc
        forms.cpython-311.pyc
        ml.cpython-311.pyc
        models.cpython-311.pyc
        routes.cpython-311.pyc
        __init__.cpython-311.pyc
```

## Routes



1. Home Page

- **Route:** `/`
- **Description:** Renders the home page of the application.

2. History Page

- **Route:** `/history`
- **Description:** Displays the user's historical car data if logged in; otherwise, redirects to the login page.

3. Wishlist Page

- **Route:** `/wishlist`
- **Description:** Displays the user's wishlist if logged in; otherwise, redirects to the login page.

4. Showcase Page

- **Route:** `/showcase`
- **Description:** Allows users to search and filter cars. Displays a catalog of cars based on the search criteria.

5. Prediction Page

- **Route:** `/predict`
- **Description:** Performs car price prediction using a regression model. If the user is logged in, adds the prediction to their history. Displays the prediction result and recommends cars.

6. Login Page

- **Route:** `/login`
- **Description:** Provides user authentication. Users can log in with their credentials.

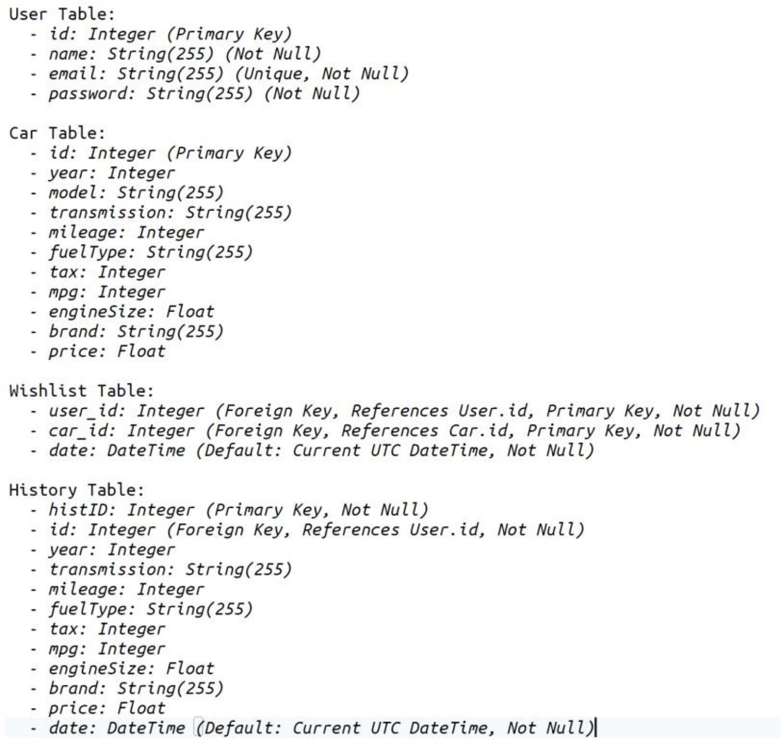
7. Signup Page

- **Route:** `/signup`
- **Description:** Allows new users to sign up for the application.

8. Profile Page

- **Route:** `/profile`
- **Description:** Displays and allows users to update their profile information if logged in; otherwise, redirects to the login page.

## Database Schema

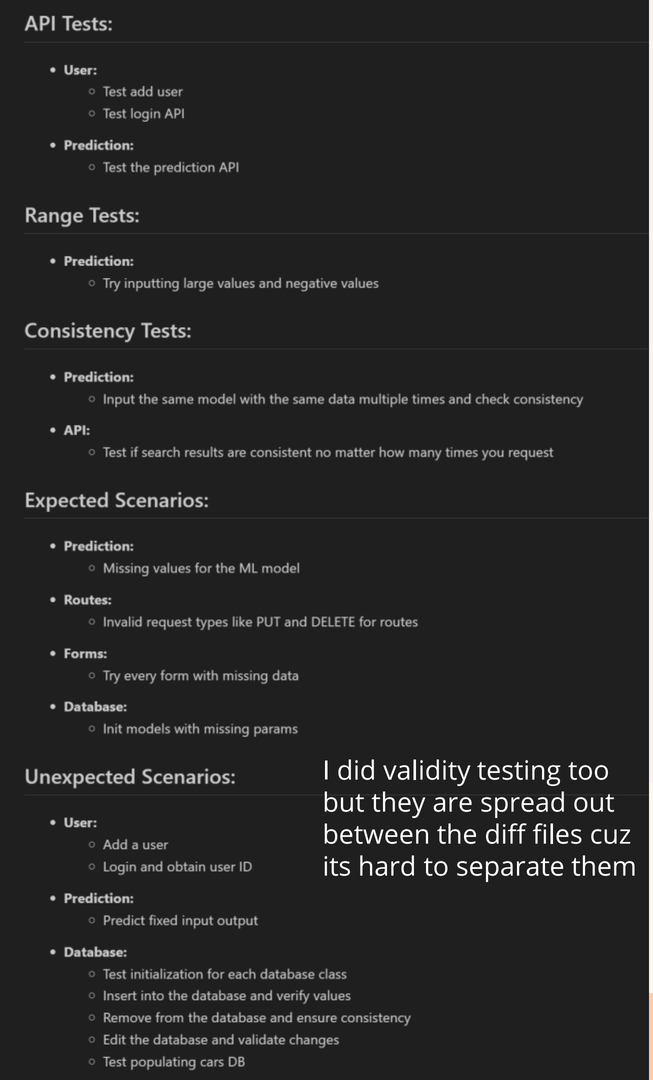


```
User Table:
  - id: Integer (Primary Key)
  - name: String(255) (Not Null)
  - email: String(255) (Unique, Not Null)
  - password: String(255) (Not Null)

Car Table:
  - id: Integer (Primary Key)
  - year: Integer
  - model: String(255)
  - transmission: String(255)
  - mileage: Integer
  - fuelType: String(255)
  - tax: Integer
  - mpg: Integer
  - engineSize: Float
  - brand: String(255)
  - price: Float

Wishlist Table:
  - user_id: Integer (Foreign Key, References User.id, Primary Key, Not Null)
  - car_id: Integer (Foreign Key, References Car.id, Primary Key, Not Null)
  - date: DateTime (Default: Current UTC DateTime, Not Null)

History Table:
  - histID: Integer (Primary Key, Not Null)
  - id: Integer (Foreign Key, References User.id, Not Null)
  - year: Integer
  - transmission: String(255)
  - mileage: Integer
  - fuelType: String(255)
  - tax: Integer
  - mpg: Integer
  - engineSize: Float
  - brand: String(255)
  - price: Float
  - date: DateTime (Default: Current UTC DateTime, Not Null)
```

# 05

# Testing

# Tests (Overview)



## API Tests:

- **User:**
  - Test add user
  - Test login API
- **Prediction:**
  - Test the prediction API

## Range Tests:

- **Prediction:**
  - Try inputting large values and negative values

## Consistency Tests:

- **Prediction:**
  - Input the same model with the same data multiple times and check consistency
- **API:**
  - Test if search results are consistent no matter how many times you request

## Expected Scenarios:

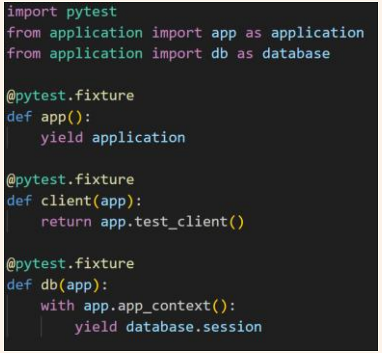
- **Prediction:**
  - Missing values for the ML model
- **Routes:**
  - Invalid request types like PUT and DELETE for routes
- **Forms:**
  - Try every form with missing data
- **Database:**
  - Init models with missing params

## Unexpected Scenarios:

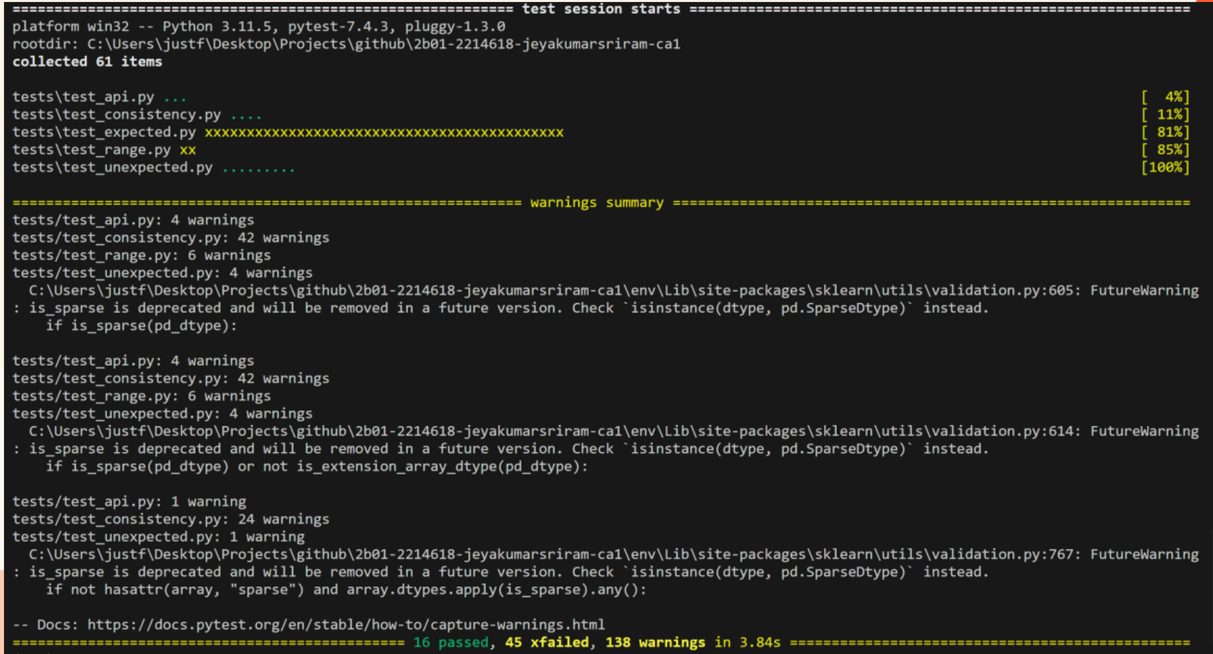
- **User:**
  - Add a user
  - Login and obtain user ID
- **Prediction:**
  - Predict fixed input output
- **Database:**
  - Test initialization for each database class
  - Insert into the database and verify values
  - Remove from the database and ensure consistency
  - Edit the database and validate changes
  - Test populating cars DB

I did validity testing too but they are spread out between the diff files cuz its hard to separate them

I used simple fixtures in my application. I group all the tests in different files based on what they are testing for like range or consistency.
All the passable tests have passed and the rest xfailed. You can ignore the warning as they are future warning about a sklearn function getting deprecated



```python
import pytest
from application import app as application
from application import db as database

@pytest.fixture
def app():
    yield application

@pytest.fixture
def client(app):
    return app.test_client()

@pytest.fixture
def db(app):
    with app.app_context():
        yield database.session
```



```
============================================================= test session starts =============================================================
platform win32 -- Python 3.11.5, pytest-7.4.3, pluggy-1.3.0
rootdir: C:\Users\justf\Desktop\Projects\github\2b01-2214618-jeyakumarsriram-ca1
collected 61 items

tests\test_api.py ...                                                                                                                   [  4%]
tests\test_consistency.py ....                                                                                                          [ 11%]
tests\test_expected.py xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx                                                                     [ 81%]
tests\test_range.py xx                                                                                                                  [ 85%]
tests\test_unexpected.py .........                                                                                                      [100%]

============================================================== warnings summary ===============================================================
tests/test_api.py: 4 warnings
tests/test_consistency.py: 42 warnings
tests/test_range.py: 6 warnings
tests/test_unexpected.py: 4 warnings
  C:\Users\justf\Desktop\Projects\github\2b01-2214618-jeyakumarsriram-ca1\env\Lib\site-packages\sklearn\utils\validation.py:605: FutureWarning
: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):

tests/test_api.py: 4 warnings
tests/test_consistency.py: 42 warnings
tests/test_range.py: 6 warnings
tests/test_unexpected.py: 4 warnings
  C:\Users\justf\Desktop\Projects\github\2b01-2214618-jeyakumarsriram-ca1\env\Lib\site-packages\sklearn\utils\validation.py:614: FutureWarning
: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype) or not is_extension_array_dtype(pd_dtype):

tests/test_api.py: 1 warning
tests/test_consistency.py: 24 warnings
tests/test_unexpected.py: 1 warning
  C:\Users\justf\Desktop\Projects\github\2b01-2214618-jeyakumarsriram-ca1\env\Lib\site-packages\sklearn\utils\validation.py:767: FutureWarning
: is_sparse is deprecated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if not hasattr(array, "sparse") and array.dtypes.apply(is_sparse).any():

-- Docs: https://docs.pytest.org/en/stable/how-to/capture-warnings.html
================================================ 16 passed, 45 xfailed, 138 warnings in 3.84s =================================================
```

# 06

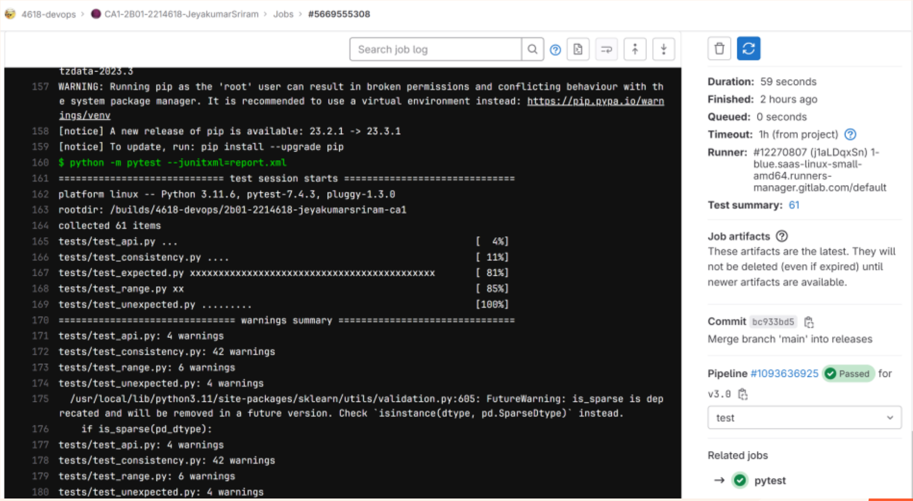
# Advanced

# Advanced

- Automatic Testing: I used gitlab's CI pipeline feature to construct one that has just one stage which is testing. Basically whenever I merge a branch or commit the tests I've created will be run and produce a report like this =>>>

- Dockerise: throughout my development I used docker to avoid touching my local environment. Using gunicorn

- Deployed to Render.com using GitLab. Initially I tried to deploy my image to dockerhub and then to render but to make things simpler I just connected my render to my releases branch to deploy every release


4618-devops › CA1-2B01-2214618-JeyakumarSriram › Jobs › #5669555308
Search job log

```
tzdata-2023.3
WARNING: Running pip as the 'root' user can result in broken permissions and conflicting behaviour with th
e system package manager. It is recommended to use a virtual environment instead: https://pip.pypa.io/warn
ings/venv
[notice] A new release of pip is available: 23.2.1 -> 23.3.1
[notice] To update, run: pip install --upgrade pip
$ python -m pytest --junitxml=report.xml
============================ test session starts ============================
platform linux -- Python 3.11.6, pytest-7.4.3, pluggy-1.3.0
rootdir: /builds/4618-devops/2b01-2214618-jeyakumarsriram-ca1
collected 61 items
tests/test_api.py ...                                                [  4%]
tests/test_consistency.py ....                                       [ 11%]
tests/test_expected.py xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx    [ 81%]
tests/test_range.py xx                                               [ 85%]
tests/test_unexpected.py .........                                   [100%]
============================== warnings summary ==============================
tests/test_api.py: 4 warnings
tests/test_consistency.py: 42 warnings
tests/test_range.py: 6 warnings
tests/test_unexpected.py: 4 warnings
  /usr/local/lib/python3.11/site-packages/sklearn/utils/validation.py:605: FutureWarning: is_sparse is dep
recated and will be removed in a future version. Check `isinstance(dtype, pd.SparseDtype)` instead.
    if is_sparse(pd_dtype):
tests/test_api.py: 4 warnings
tests/test_consistency.py: 42 warnings
tests/test_range.py: 6 warnings
tests/test_unexpected.py: 4 warnings
```

**Duration:** 59 seconds
**Finished:** 2 hours ago
**Queued:** 0 seconds
**Timeout:** 1h (from project)
**Runner:** #12270807 (j1aLDqxSn) 1-blue.saas-linux-small-amd64.runners-manager.gitlab.com/default
**Test summary:** 61

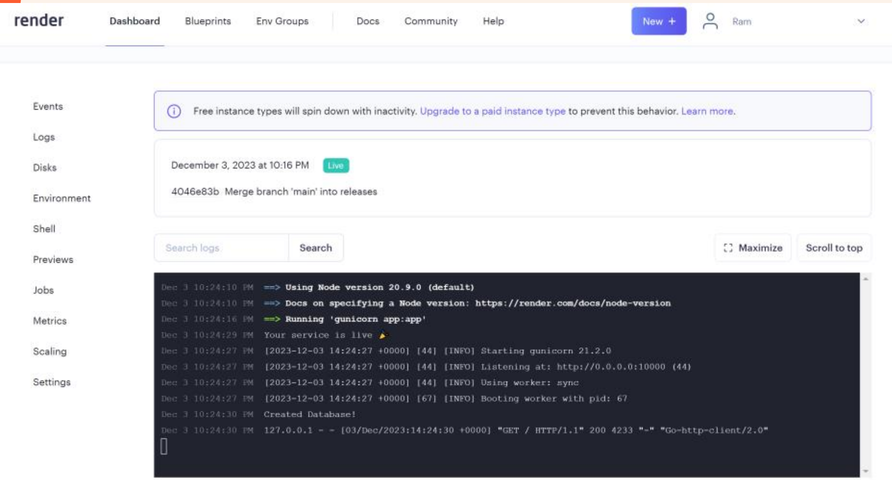
**Job artifacts**
These artifacts are the latest. They will not be deleted (even if expired) until newer artifacts are available.

**Commit** bc933bd5
Merge branch 'main' into releases

**Pipeline** #1093636925 Passed for v3.0

test

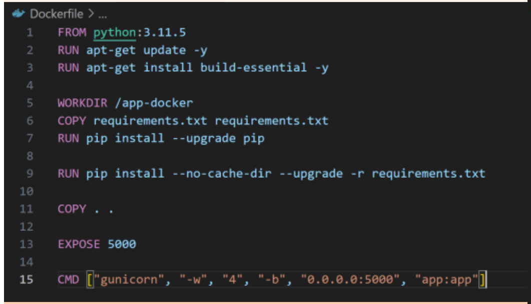
Related jobs
→ pytest



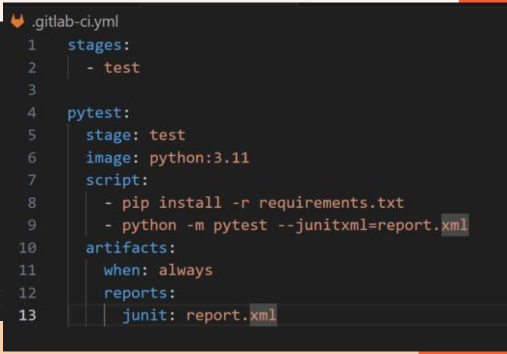
Free instance types will spin down with inactivity. Upgrade to a paid instance type to prevent this behavior. Learn more.

December 3, 2023 at 10:16 PM Live
4046e83b Merge branch 'main' into releases

```
Dec 3 10:24:10 PM ==> Using Node version 20.9.0 (default)
Dec 3 10:24:10 PM ==> Docs on specifying a Node version: https://render.com/docs/node-version
Dec 3 10:24:16 PM ==> Running 'gunicorn app:app'
Dec 3 10:24:29 PM Your service is live 🎉
Dec 3 10:24:27 PM [2023-12-03 14:24:27 +0000] [44] [INFO] Starting gunicorn 21.2.0
Dec 3 10:24:27 PM [2023-12-03 14:24:27 +0000] [44] [INFO] Listening at: http://0.0.0.0:10000 (44)
Dec 3 10:24:27 PM [2023-12-03 14:24:27 +0000] [44] [INFO] Using worker: sync
Dec 3 10:24:27 PM [2023-12-03 14:24:27 +0000] [67] [INFO] Booting worker with pid: 67
Dec 3 10:24:30 PM Created Database!
Dec 3 10:24:30 PM 127.0.0.1 - - [03/Dec/2023:14:24:30 +0000] "GET / HTTP/1.1" 200 4233 "-" "Go-http-client/2.0"
```



```dockerfile
FROM python:3.11.5
RUN apt-get update -y
RUN apt-get install build-essential -y

WORKDIR /app-docker
COPY requirements.txt requirements.txt
RUN pip install --upgrade pip

RUN pip install --no-cache-dir --upgrade -r requirements.txt

COPY . .

EXPOSE 5000

CMD ["gunicorn", "-w", "4", "-b", "0.0.0.0:5000", "app:app"]
```



```yaml
stages:
  - test

pytest:
  stage: test
  image: python:3.11
  script:
    - pip install -r requirements.txt
    - python -m pytest --junitxml=report.xml
  artifacts:
    when: always
    reports:
      junit: report.xml
```

# Thanks!

## Do you have any questions?

youremail@freepik.com
+34 654 321 432
yourwebsite.com

**CREDITS:** This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution