In [1]:

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

In [2]:

```
#import libraries
import requests
import pandas as pd
import numpy as np
import csv
import urllib.request
from bs4 import BeautifulSoup
```

In [3]:

```
import random
file_path = '/content/drive/My Drive/data_scrap_oyunerdene/car1.tsv'
with open(file_path, 'r', encoding='utf-8') as f:
    app_lines = f.read().split('\n')
```

In [4]:

```
df = pd.read_csv('/content/drive/My Drive/data_scrap_oyunerdene/car1.tsv' , sep='\t')
```

In [5]:

```
len(df)
```

Out[5]:

```
360
```

In [6]:

```
df.head(10)
```

Out[6]:

	une	motor	hurdnii_hairtsag	hurd	turul	ungu	uild_on	orj_on	hudulguur	dotor_ungu	leasing	hutlugchn
0	**480.0**	**4.0**	Автомат	Зөв	Жийп	Боронзон	**2021**	**2021**	Бензин	Хар	Лизинггүй	Бүх дугуй **4WD**
1	**45.0**	**3.2**	Автомат	Зөв	Суудлын тэрэг	Хар	**1994**	**2024**	Бензин	Хар	Лизинггүй	Хойноо **RWD**
2	**65.0**	**5.0**	Автомат	Зөв	Жийп	Хар	**2002**	**2012**	Бензин	Хар	Лизинггүй	Бүх дугуй **4WD**
3	**45.0**	**3.3**	Автомат	Зөв	Суудлын тэрэг	Цагаан	**2014**	**2018**	Бензин	Хар	Лизинггүй	Бүх дугуй **4WD**
4	**47.0**	**1.6**	Автомат	Буруу	Суудлын тэрэг	Хар	**2018**	**2020**	Бензин	Хар	Лизинггүй	Урдаа **FWD**
5	**150.0**	**2.0**	Автомат	Буруу	Гэр бүлийн	Бусад	**2016**	**2023**	Дизель	Бусад	Лизинггүй	Бүх дугуй **4WD**
6	**29.9**	**1.6**	Автомат	Буруу	Суудлын тэрэг	Бор	**2013**	**2023**	Бензин	Хар	Лизинггүй	Урдаа **FWD**
7	**95.0**	**5.5**	Автомат	Зөв	Жийп	Хүрэн	**2009**	**2009**	Бензин	Бусад	ББСБ лизингтэй	Бүх дугуй **4WD**
8	**85.0**	**0.6**	Автомат	Буруу	Жийп	Цайвар цэнхэр	**2017**	**2022**	Бензин	Хар	Лизинггүй	Бүх дугуй **4WD**
9	**160.0**	**3.0**	Автомат	Зөв	Суудлын тэрэг	Хар	**2016**	**2024**	Бензин	Хар	Банкны лизингтэй	Бүх дугуй **4WD**

une motor hurdnii_hairtsag hurd turul ungu uild_on orj_on hudulguur dotor_ungu leasing hutlugchn

In [7]:

```
filtered_df = df[df.hudulguur == 'Бензин']

print(len(df),len(filtered_df))

for index, row in filtered_df.iterrows():
  if ('Автомат' in row['hurdnii_hairtsag']) is False:
    print(row)
```

```
360 336
```

In [8]:

```
import pandas as pd

def normalizeDataSet(app_set):
  ret = pd.DataFrame(columns=["une ","motor",'hurdnii_hairtsag','hurd','turul','ungu','u
ild_on', 'orj_on', 'hudulguur', 'dotor_ungu', 'leasing', 'hutlugchn', 'kilo', 'nuhtsul',
'haalga'])
  for index, row in app_set.iterrows():


    #print(float(row['space']))
    if row['leasing'].strip()=='Лизингтүй':
      leas = False
    else:
      leas = True

    ret = ret.append({'une':row['une'] , 'motor': row['motor'],'hurdnii_hairtsag': row['
hurdnii_hairtsag'], 'hurd': row['hurd'], 'turul' : row['turul'],
                  'ungu': row['ungu'],'uild_on': row['uild_on'], 'orj_on': row['orj_on']
,'hudulguur': row['hudulguur'],'dotor_ungu': row['dotor_ungu'], 'leasing': row['leasing'
],
                  'hutlugchn': row['hutlugchn'], 'kilo': row['kilo'],'nuhtsul': row['nuh
tsul'], 'haalga': row['haalga']}, ignore_index=True)
  return ret
```

In [9]:

```
print(filtered_df.columns)
```

```
Index(['une', 'motor', 'hurdnii_hairtsag', 'hurd', 'turul', 'ungu', 'uild_on',
       'orj_on', 'hudulguur', 'dotor_ungu', 'leasing', 'hutlugchn', 'kilo',
       'nuhtsul', 'haalga'],
      dtype='object')
```

In [14]:

```
for col in app_df.columns:
    if col not in filtered_df.columns:
        print(f"Column '{col}' not found in filtered_df")
```

In [12]:

```
app_df = pd.DataFrame(columns=['une', 'motor', 'hurdnii_hairtsag', 'hurd', 'turul', 'ung
u', 'uild_on', 'orj_on', 'hudulguur', 'dotor_ungu', 'leasing', 'hutlugchn', 'kilo', 'nuh
tsul', 'haalga'])

for index, row in filtered_df.iterrows():
    app_df.loc[index] = {'une': row['une'], 'motor': row['motor'], 'hurdnii_hairtsag': r
ow['hurdnii_hairtsag'], 'hurd': row['hurd'], 'turul': row['turul'],
                         'ungu': row['ungu'], 'uild_on': row['uild_on'], 'orj_on': row['or
j_on'], 'hudulguur': row['hudulguur'], 'dotor_ungu': row['dotor_ungu'], 'leasing': row['
leasing'],
                         'hutlugchn': row['hutlugchn'], 'kilo': row['kilo'], 'nuhtsul': ro
w['nuhtsul'], 'haalga': row['haalga']}
```

In [13]:

In [13]:

```
app_df = pd.DataFrame(columns=filtered_df.columns)

for index, row in filtered_df.iterrows():
    app_df.loc[index] = row
```

In [15]:

```
len(app_df)
```

Out[15]:

336

In [16]:

```
app_df.to_csv('data.csv')
```

In [17]:

```
app_df.head()
```

Out[17]:

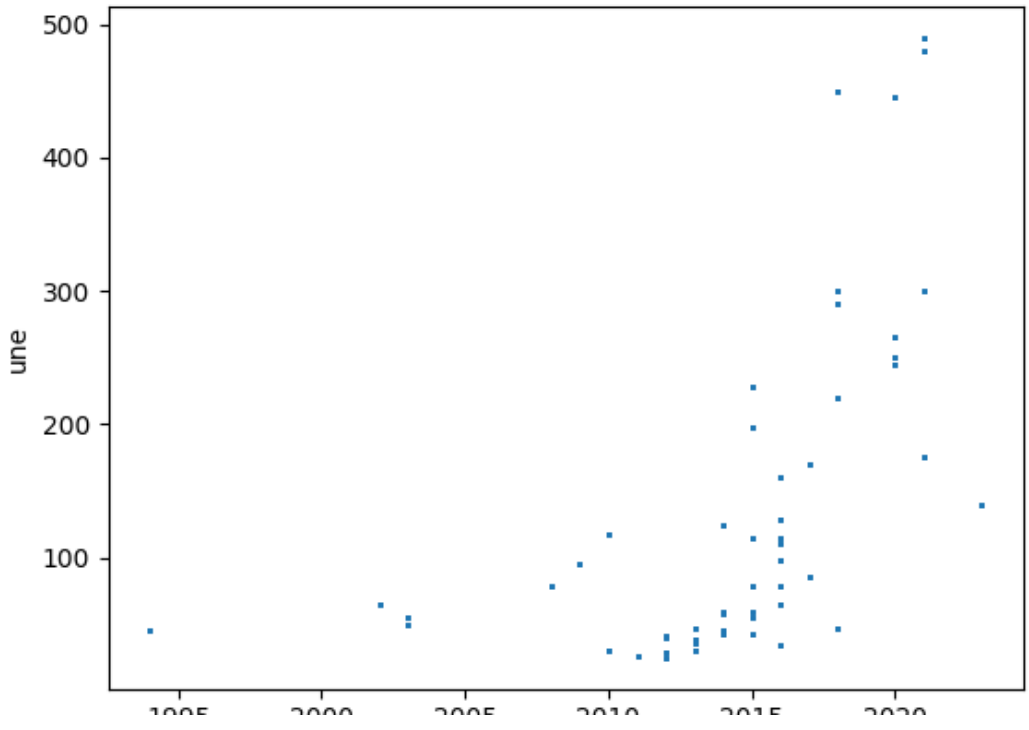
	une	motor	hurdnii_hairtsag	hurd	turul	ungu	uild_on	orj_on	hudulguur	dotor_ungu	leasing	hutlugchn
0	**480.0**	**4.0**	Автомат	Зөв	Жийп	Боронзон	**2021**	**2021**	Бензин	Хар	Лизинггүй	Бүх дугуй **4WD**
1	**45.0**	**3.2**	Автомат	Зөв	Суудлын тэрэг	Хар	**1994**	**2024**	Бензин	Хар	Лизинггүй	Хойноо **RWD**
2	**65.0**	**5.0**	Автомат	Зөв	Жийп	Хар	**2002**	**2012**	Бензин	Хар	Лизинггүй	Бүх дугуй **4WD**
3	**45.0**	**3.3**	Автомат	Зөв	Суудлын тэрэг	Цагаан	**2014**	**2018**	Бензин	Хар	Лизинггүй	Бүх дугуй **4WD**
4	**47.0**	**1.6**	Автомат	Буруу	Суудлын тэрэг	Хар	**2018**	**2020**	Бензин	Хар	Лизинггүй	Урдаа **FWD**

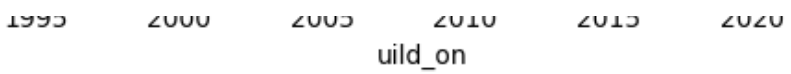
In [18]:

```
app_df.plot.scatter(x = 'uild_on', y = 'une', s = 1)
```

Out[18]:

<Axes: xlabel='uild_on', ylabel='une'>



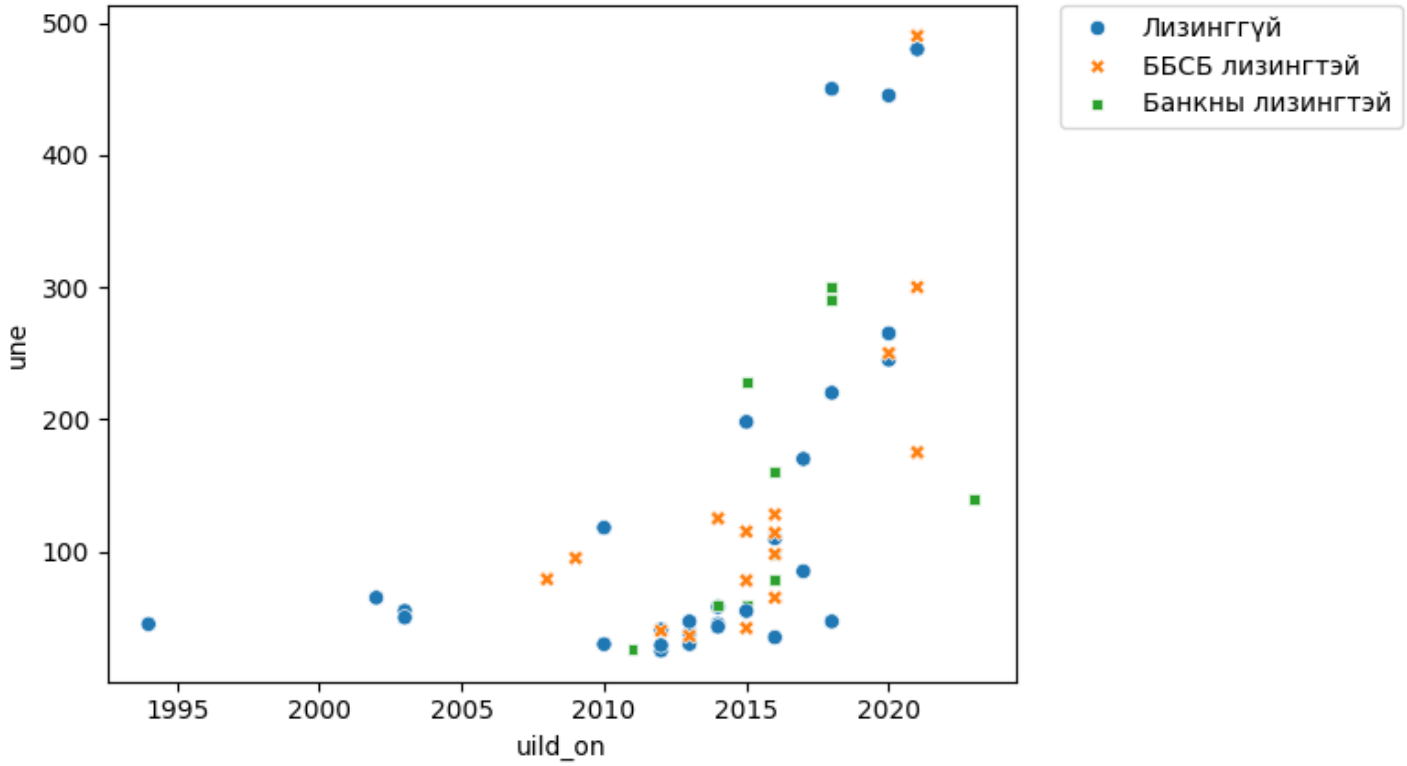


In [19]:

```
import matplotlib.pyplot as plt
import seaborn as sns
sns.scatterplot(data=app_df, x="uild_on", y="une", hue="leasing",style="leasing")
plt.legend(bbox_to_anchor=(1.05, 1), loc=2, borderaxespad=0.)
```

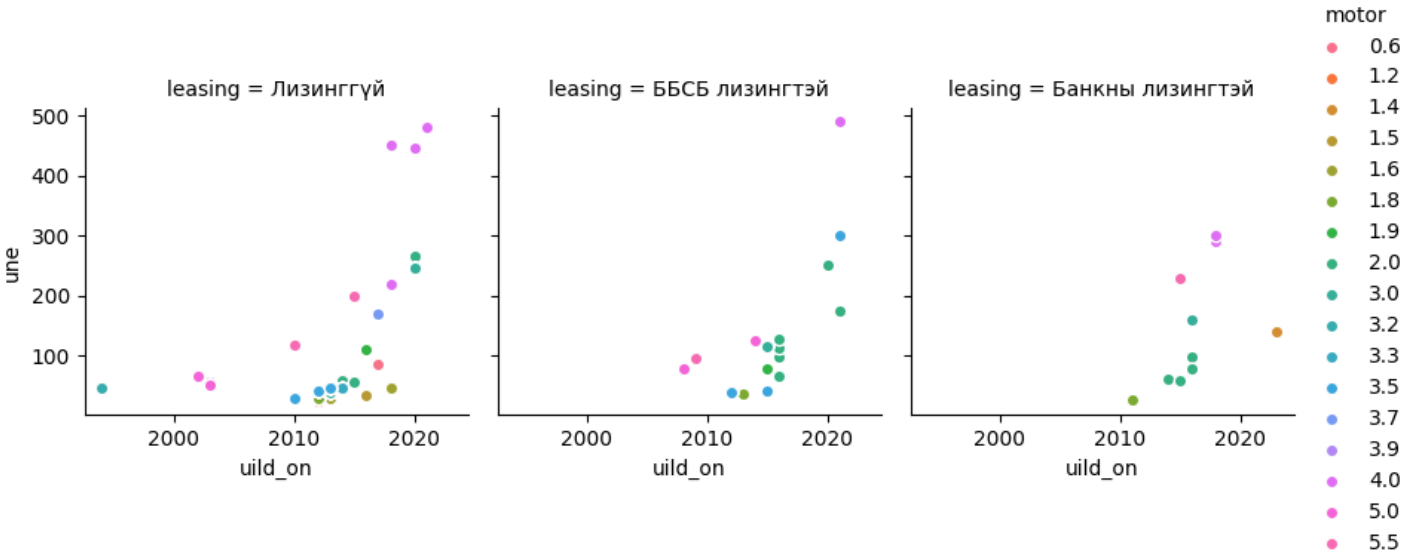
Out[19]:

```
<matplotlib.legend.Legend at 0x7bb9b4c01810>
```



In [20]:

```
graph = sns.FacetGrid(app_df, col ="leasing",  hue ="motor")
# map the above form facetgrid with some attributes
graph.map(plt.scatter, "uild_on", "une", edgecolor ="w").add_legend()
# show the object
plt.show()
```



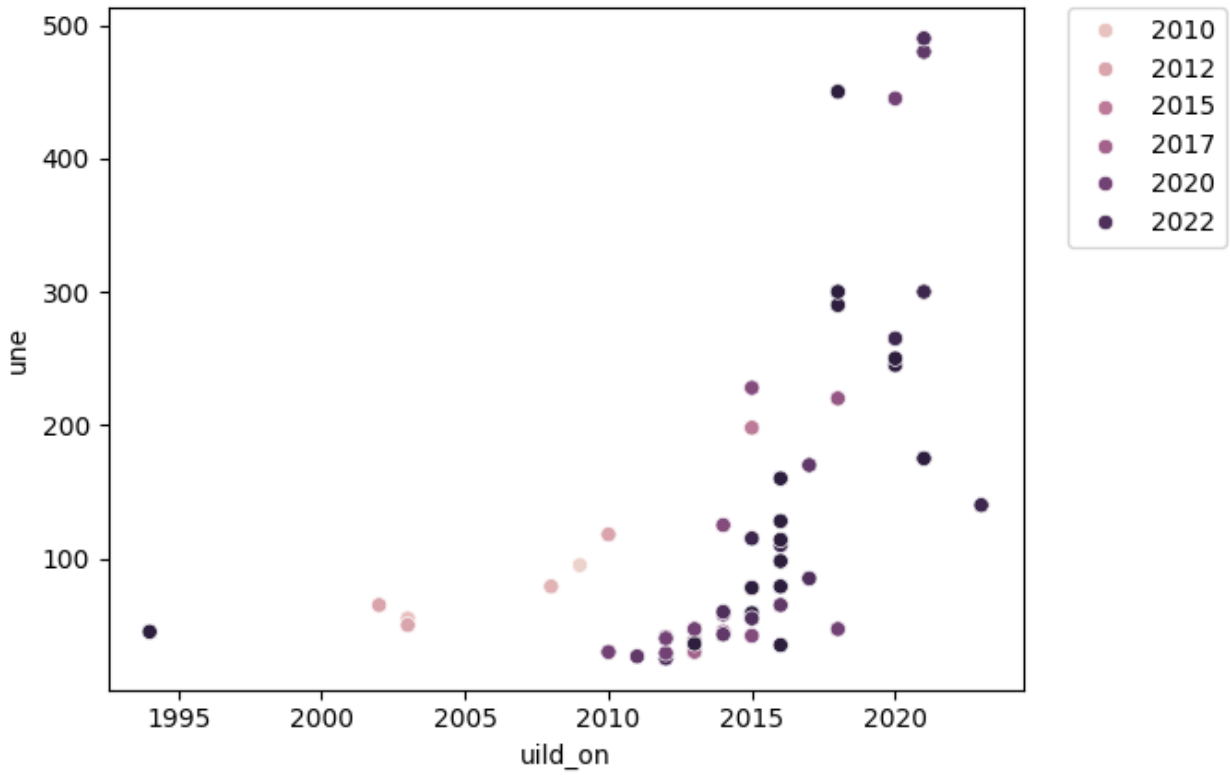
In [46]:

```
sns.scatterplot(data=app_df, x="uild_on", y="une", hue="orj_on")
```

```
plt.legend(bbox_to_anchor=(1.05, 1), loc=2, borderaxespad=0.)
```

Out[46]:

```
<matplotlib.legend.Legend at 0x7bb8cb85e410>
```

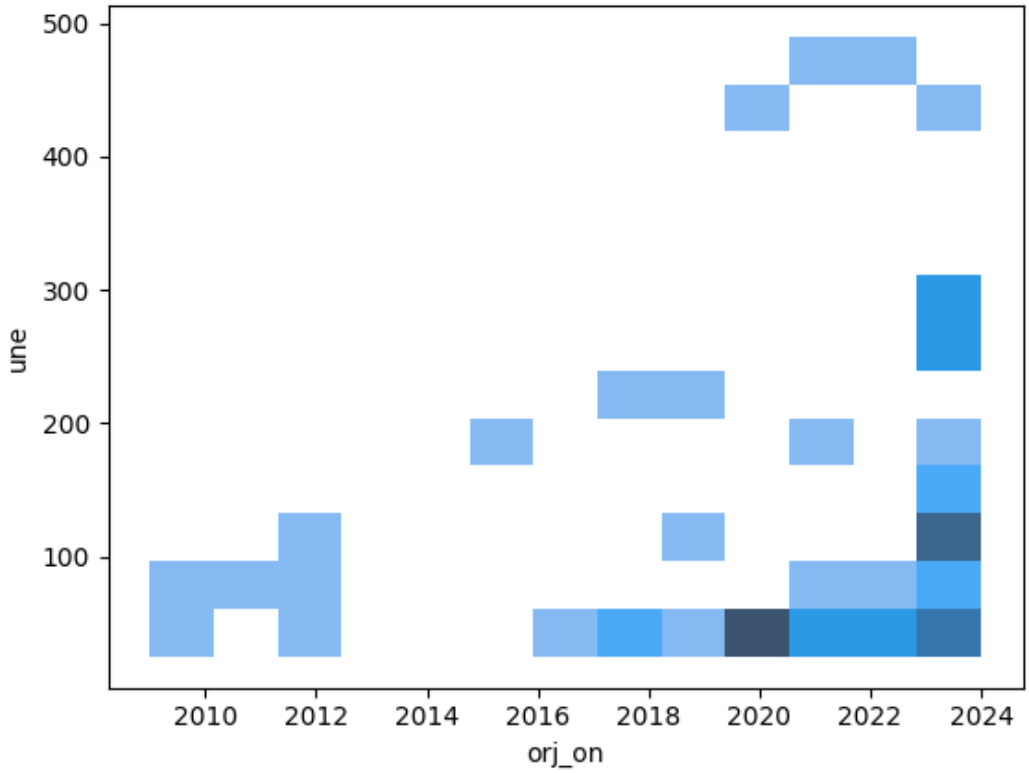


In [48]:

```
sns.histplot(data=app_df, y="une", x="orj_on")
```

Out[48]:

```
<Axes: xlabel='orj_on', ylabel='une'>
```



In [23]:

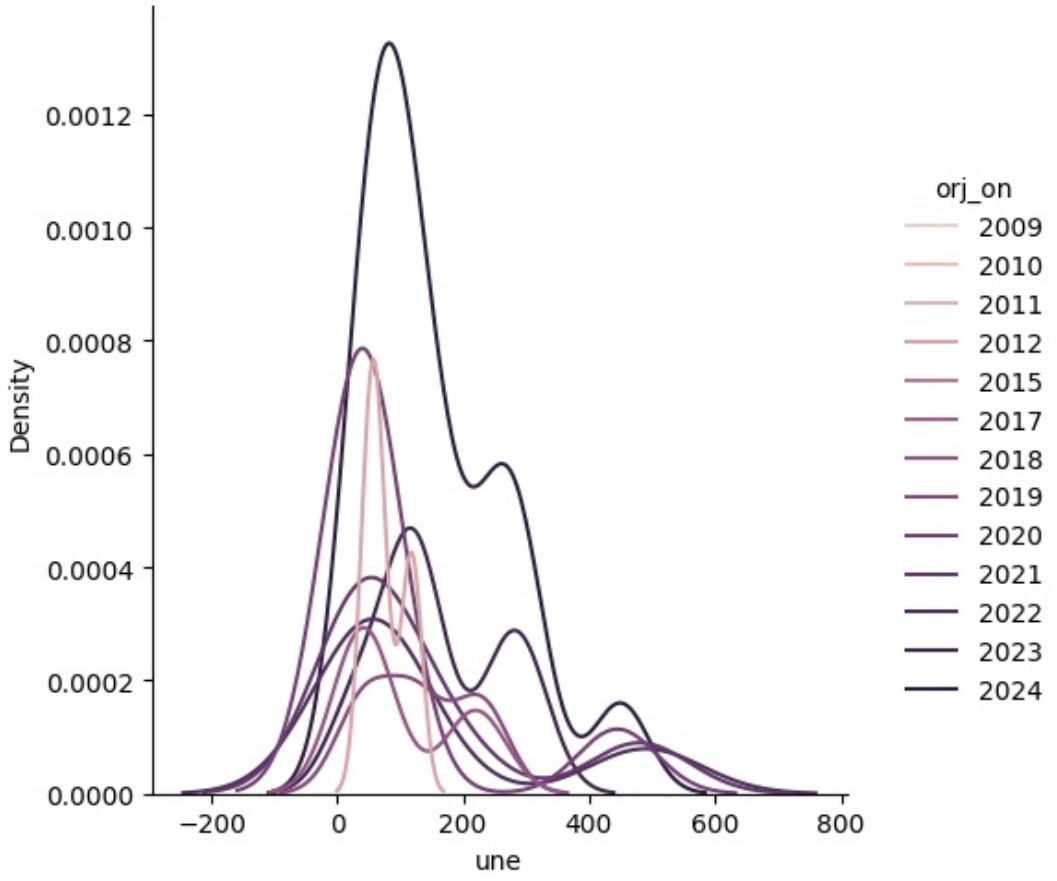
```
sns.displot(data=app_df, x="une", hue="orj_on", kind='kde')
```

```
<ipython-input-23-2178c2ded6b9>:1: UserWarning: Dataset has 0 variance; skipping density estimate. Pass `warn_singular=False` to disable this warning.
```

```
  sns.displot(data=app_df, x="une", hue="orj_on", kind='kde')
```

Out[23]:

```
<seaborn.axisgrid.FacetGrid at 0x7bb9a22c2740>
```



In [24]:

```
!pip install -U sentence-transformers
```

```
Collecting sentence-transformers
  Downloading sentence_transformers-2.6.1-py3-none-any.whl (163 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 163.3/163.3 kB 3.4 MB/s eta 0:00:00
Requirement already satisfied: transformers<5.0.0,>=4.32.0 in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (4.38.2)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (4.66.2)
Requirement already satisfied: torch>=1.11.0 in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (2.2.1+cu121)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (1.25.2)
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (1.2.2)
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (1.11.4)
Requirement already satisfied: huggingface-hub>=0.15.1 in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (0.20.3)
Requirement already satisfied: Pillow in /usr/local/lib/python3.10/dist-packages (from sentence-transformers) (9.4.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.1->sentence-transformers) (3.13.3)
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.1->sentence-transformers) (2023.6.0)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.1->sentence-transformers) (2.31.0)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.1->sentence-transformers) (6.0.1)
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.1->sentence-transformers) (4.10.0)
Requirement already satisfied: packaging>=20.9 in /usr/local/lib/python3.10/dist-packages (from huggingface-hub>=0.15.1->sentence-transformers) (24.0)
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages (from tor
```

```
ch>=1.11.0->sentence-transformers) (1.12)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from
torch>=1.11.0->sentence-transformers) (3.2.1)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages (from to
rch>=1.11.0->sentence-transformers) (3.1.3)
Collecting nvidia-cuda-nvrtc-cu12==12.1.105 (from torch>=1.11.0->sentence-transformers)
  Using cached nvidia_cuda_nvrtc_cu12-12.1.105-py3-none-manylinux1_x86_64.whl (23.7 MB)
Collecting nvidia-cuda-runtime-cu12==12.1.105 (from torch>=1.11.0->sentence-transformers)
  Using cached nvidia_cuda_runtime_cu12-12.1.105-py3-none-manylinux1_x86_64.whl (823 kB)
Collecting nvidia-cuda-cupti-cu12==12.1.105 (from torch>=1.11.0->sentence-transformers)
  Using cached nvidia_cuda_cupti_cu12-12.1.105-py3-none-manylinux1_x86_64.whl (14.1 MB)
Collecting nvidia-cudnn-cu12==8.9.2.26 (from torch>=1.11.0->sentence-transformers)
  Using cached nvidia_cudnn_cu12-8.9.2.26-py3-none-manylinux1_x86_64.whl (731.7 MB)
Collecting nvidia-cublas-cu12==12.1.3.1 (from torch>=1.11.0->sentence-transformers)
  Using cached nvidia_cublas_cu12-12.1.3.1-py3-none-manylinux1_x86_64.whl (410.6 MB)
Collecting nvidia-cufft-cu12==11.0.2.54 (from torch>=1.11.0->sentence-transformers)
  Using cached nvidia_cufft_cu12-11.0.2.54-py3-none-manylinux1_x86_64.whl (121.6 MB)
Collecting nvidia-curand-cu12==10.3.2.106 (from torch>=1.11.0->sentence-transformers)
  Using cached nvidia_curand_cu12-10.3.2.106-py3-none-manylinux1_x86_64.whl (56.5 MB)
Collecting nvidia-cusolver-cu12==11.4.5.107 (from torch>=1.11.0->sentence-transformers)
  Using cached nvidia_cusolver_cu12-11.4.5.107-py3-none-manylinux1_x86_64.whl (124.2 MB)
Collecting nvidia-cusparse-cu12==12.1.0.106 (from torch>=1.11.0->sentence-transformers)
  Using cached nvidia_cusparse_cu12-12.1.0.106-py3-none-manylinux1_x86_64.whl (196.0 MB)
Collecting nvidia-nccl-cu12==2.19.3 (from torch>=1.11.0->sentence-transformers)
  Using cached nvidia_nccl_cu12-2.19.3-py3-none-manylinux1_x86_64.whl (166.0 MB)
Collecting nvidia-nvtx-cu12==12.1.105 (from torch>=1.11.0->sentence-transformers)
  Using cached nvidia_nvtx_cu12-12.1.105-py3-none-manylinux1_x86_64.whl (99 kB)
Requirement already satisfied: triton==2.2.0 in /usr/local/lib/python3.10/dist-packages (
from torch>=1.11.0->sentence-transformers) (2.2.0)
Collecting nvidia-nvjitlink-cu12 (from nvidia-cusolver-cu12==11.4.5.107->torch>=1.11.0->s
entence-transformers)
  Using cached nvidia_nvjitlink_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (21.1 MB)
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packag
es (from transformers<5.0.0,>=4.32.0->sentence-transformers) (2023.12.25)
Requirement already satisfied: tokenizers<0.19,>=0.14 in /usr/local/lib/python3.10/dist-p
ackages (from transformers<5.0.0,>=4.32.0->sentence-transformers) (0.15.2)
Requirement already satisfied: safetensors>=0.4.1 in /usr/local/lib/python3.10/dist-packa
ges (from transformers<5.0.0,>=4.32.0->sentence-transformers) (0.4.2)
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.10/dist-packages (
from scikit-learn->sentence-transformers) (1.3.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-pac
kages (from scikit-learn->sentence-transformers) (3.4.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages
(from jinja2->torch>=1.11.0->sentence-transformers) (2.1.5)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist
-packages (from requests->huggingface-hub>=0.15.1->sentence-transformers) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (f
rom requests->huggingface-hub>=0.15.1->sentence-transformers) (3.6)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packa
ges (from requests->huggingface-hub>=0.15.1->sentence-transformers) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packa
ges (from requests->huggingface-hub>=0.15.1->sentence-transformers) (2024.2.2)
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-packages (f
rom sympy->torch>=1.11.0->sentence-transformers) (1.3.0)
Installing collected packages: nvidia-nvtx-cu12, nvidia-nvjitlink-cu12, nvidia-nccl-cu12,
nvidia-curand-cu12, nvidia-cufft-cu12, nvidia-cuda-runtime-cu12, nvidia-cuda-nvrtc-cu12,
nvidia-cuda-cupti-cu12, nvidia-cublas-cu12, nvidia-cusparse-cu12, nvidia-cudnn-cu12, nvid
ia-cusolver-cu12, sentence-transformers
Successfully installed nvidia-cublas-cu12-12.1.3.1 nvidia-cuda-cupti-cu12-12.1.105 nvidia
-cuda-nvrtc-cu12-12.1.105 nvidia-cuda-runtime-cu12-12.1.105 nvidia-cudnn-cu12-8.9.2.26 nv
idia-cufft-cu12-11.0.2.54 nvidia-curand-cu12-10.3.2.106 nvidia-cusolver-cu12-11.4.5.107 n
vidia-cusparse-cu12-12.1.0.106 nvidia-nccl-cu12-2.19.3 nvidia-nvjitlink-cu12-12.4.127 nvi
dia-nvtx-cu12-12.1.105 sentence-transformers-2.6.1
```

In [25]:

```
from sentence_transformers import SentenceTransformer
sen_model = SentenceTransformer('sentence-transformers/all-mpnet-base-v2')
```

```
/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/_token.py:88: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://h
```

```
uggingface.co/settings/tokens), set it as secret in your Google Colab and restart your se
ssion.
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models
or datasets.
  warnings.warn(
```

In [30]:

```
sentences = []
y = []
z = []
for index, row in app_df.iterrows():
  sentences.append(row['turul']+' '+row['hutlugchn'])
  z.append(row['une'])
  y.append(row['orj_on'])
```

In [31]:

```
sentence_vectors = sen_model.encode(sentences)
```

In [32]:

```
it = 0
x_ = []
for i in range(len(sentence_vectors)):
  B = np.array([z[it]])
  # B 1 x 1
  # Concat sentence vector with m2 value
  O = np.append(sentence_vectors[i], B)
  # O 1 x 100 + 1 x 1 = 1 x 101
  x_.append(O)
  it+=1

# x_ 5000 x 101
# y  5000 x 1
```

In [33]:

```
from sklearn.linear_model import LinearRegression
model = LinearRegression(fit_intercept=True)

model.fit(x_, y)
```

Out[33]:

▾ LinearRegression

LinearRegression()

In [41]:

```
def extractSpace(sentence):
  ans = int(sum(z)/len(z))
  if 'Суудлын тэрэг' in sentence or 'Жиип' in sentence:
    sentence = sentence.replace('Суудлын тэрэг','Жиип')
    sentence = sentence.replace('Жиип','Жиип')
    tokens = sentence.split(' ')
    for token in tokens:
      if token.endswith(' ') and len(token) > 2:
          ans = int(token[:len(token)-2])
          return ans
  return ans
```

```
def predictEstimate(sentence):
  p_embeddings = sen_model.encode([sentence])
  d = [extractSpace(sentence)]
  p_ = []
  for emb in p_embeddings:
    B = [d[0]]
    p_.append(np.append( emb , B ))
  y_prediction =  model.predict(p_)
  return y_prediction[0]
```

In [43]:

```
print(predictEstimate(" Жиип 2015 Бүх дугуй 4WD"))
```

```
171754.76138329564
```