

Міністерство освіти і науки України
Національний університет «Львівська політехніка»
Інститут комп'ютерних наук та інформаційних технологій
Кафедра «Системи штучного інтелекту»



Лабораторна робота №1
з курсу «Аналітичні сховища даних»

Виконала:

студентка групи КН-308

Ріжко Марія

Перевірила:

Кривенчук Ю.П.

Львів 2021 р.

Тема:

Створення і наповнення сховища даних

Завдання:

1. Відповідно до заданого варіанту підготувати необхідні дані у вигляді таблиць MS Excel і зберегти їх як персональні файли. Для підготовки даних використовувати тематичні сайти Інтернет, результати проходження практик, довідники і каталоги.

2. Створити персональне сховище даних і організувати доступ до нього відповідно до методики, представленої в розділі «Порядок виконання роботи. Частина 1».

3. Виконати завантаження даних з таблиць MS Excel за допомогою Майстра імпорту відповідно до методики, представленої в розділі «Порядок виконання роботи. Частина 2».

4. Організувати завантаження даних за допомогою Майстра експорту в сховище даних відповідно до методики, представленої в розділі «Порядок виконання роботи. Частина 3».

5. Здійснити наступний імпорт даних з сховища: кількість відвантаженого товару в розрізі дат та товарів по вибраному Вами клієнтові, залишивши одну властивість товару (вибір властивості довільний). Виконання завдання здійснюється відповідно до методики, представленої в розділі «Порядок виконання роботи. Частина 4».

1. Створити OLAP–куб і отримати крос-таблицю для варіанту сховища даних, розробленого в попередній лабораторній роботі відповідно до заданого варіанту.

2. Виконати всі операції з крос-таблицею, відповідно до методики, представленої в розділі «Порядок виконання роботи. Частина 1».

3. Здійснити агрегацію даних відповідно до методики, представленої в розділі «Порядок виконання роботи. Частина 2».

4. Виконати фільтрацію даних крос-таблиці відповідно до методики, представленої в розділі «Порядок виконання роботи. Частина 3».

Результат

Підготувати дані.

	A	B	C	D	E	F	G	H	I	J	K
1	buyer	car	price	body	mileage	engV	engType	registration	year	model	drive
2	1404	Ford	15500.00	crossover	68	2.5	Gas	yes	2010	Kuga	full
3	1502	Mercedes-Benz	20500.00	sedan	173	1.8	Gas	yes	2011	E-Class	rear
4	1084	Mercedes-Benz	35000.00	other	135	5.5	Petrol	yes	2008	CL 550	rear
5	1699	Mercedes-Benz	17800.00	van	162	1.8	Diesel	yes	2012	B 180	front
6	342	Mercedes-Benz	33000.00	vagon	91	NA	Other	yes	2013	E-Class	
7	434	Nissan	16600.00	crossover	83	2	Petrol	yes	2013	X-Trail	full
8	1762	Honda	6500.00	sedan	199	2	Petrol	yes	2003	Accord	front
9	1106	Renault	10500.00	vagon	185	1.5	Diesel	yes	2011	Megane	front
10	1593	Mercedes-Benz	21500.00	sedan	146	1.8	Gas	yes	2012	E-Class	rear
11	1002	Mercedes-Benz	22700.00	sedan	125	2.2	Diesel	yes	2010	E-Class	rear
12	1327	Nissan	20447.15	crossover	0	1.2	Petrol	yes	2016	Qashqai	front
13	629	Mercedes-Benz	20400.00	sedan	190	1.8	Gas	yes	2011	E-Class	rear
14	1791	Mercedes-Benz	22500.00	sedan	164	1.8	Gas	yes	2012	E-Class	rear
15	607	BMW	4700.00	sedan	200	NA	Petrol	yes	1996	316	rear
16	1866	Mercedes-Benz	21500.00	sedan	159	1.8	Gas	yes	2012	E-Class	rear
17	1383	BMW	19999.00	sedan	290	4.8	Petrol	yes	2006	750	rear
18	615	BMW	129222.00	sedan	2	5	Petrol	yes	2016	750	full
19	1296	Mercedes-Benz	99999.00	crossover	0	3	Petrol	yes	2016	GLE-Class	full
20	1713	Nissan	16600.00	crossover	83	2	Petrol	yes	2013	X-Trail	full
21	1447	BMW	73900.00	sedan	57	4.4	Petrol	yes	2013	M5	rear
22	19	Land Rover	0.00	crossover	0	4.4	Diesel	yes	2016	Range Rover	full
23	28	Nissan	26033.55	crossover	0	1.6	Diesel	yes	2016	X-Trail	full
24	1173	BMW	104999.00	crossover	2	3	Diesel	yes	2016	X5	full
25	1005	BMW	66500.00	crossover	1	2	Diesel	yes	2016	X5	full

Завантажити дані.

[2]:

```
import pandas as pd
```

[3]:

```
df = pd.read_excel('cars.xlsx')
df.head()
```

[3]:

	buyer	car	price	body	mileage	engV	engType	registration	year	model	drive
0	1964	Ford	15500.0	crossover	68	2.5	Gas	yes	2010	Kuga	full
1	460	Mercedes-Benz	20500.0	sedan	173	1.8	Gas	yes	2011	E-Class	rear
2	1283	Mercedes-Benz	35000.0	other	135	5.5	Petrol	yes	2008	CL 550	rear
3	604	Mercedes-Benz	17800.0	van	162	1.8	Diesel	yes	2012	B 180	front
4	298	Mercedes-Benz	33000.0	vagon	91	NaN	Other	yes	2013	E-Class	NaN

Вибрати всі записи для конкретного клієнта.

[4]:

df[df['buyer']==604]

[4]:

	buyer	car	price	body	mileage	engV	engType	registration	year	model	drive	
	3	604	Mercedes-Benz	17800.0	van	162	1.8	Diesel	yes	2012	B 180	front
	264	604	Honda	12990.0	sedan	111	1.8	Petrol	yes	2011	Civic	front
	6487	604	Geely	8300.0	sedan	33	1.8	Gas	yes	2013	Emgrand 7 (EC7)	NaN
	8062	604	Seat	7900.0	hatch	186	2.0	Petrol	yes	2007	Leon	front
	8727	604	GAZ	2000.0	sedan	70	70.0	Petrol	yes	1968	21	rear

Отримати крос таблицю.

```
[31]: df1 = df[df['year'] > 2000]
pd.pivot_table(df1, values='price', index=['buyer'], columns=['year'], aggfunc=np.sum, fill_value='')
```

```
[31]: year 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016
buyer
0          7500 6900          21900
1  4700          9500          47000
2          7300  9850  5550 16200          15800
3          4500 50700  0  4300 37000          58500
4          10500 6700
...
1996          19500 12350          11900
1997          40669
1998          14500          41500
1999          16555          36000
2000          4700  11000          6261.45 22800
```

1962 rows x 16 columns

Використати інші методи агрегації мін / макс / середнє / кількість.

Мін.

```
[32]: pd.pivot_table(df1, values='price', index=['buyer'], columns=['year'], aggfunc=np.min, fill_value='')
```

```
[32]: year 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016
buyer
0          7500 6900          21900
1  4700          9500          47000
2          7300  9850  5550 16200          15800
3          4500 13200  0  4300 14200          58500
4          10500 6700
...
1996          19500 12350          11900
1997          7170
1998          14500          41500
1999          16555          15500
2000          4700  11000          6261.45 22800
```

Макс.

```
[33]: pd.pivot_table(df1, values='price', index=['buyer'], columns=['year'], aggfunc=np.max, fill_value='')
```

```
[33]:
```

	year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
buyer																	
0				7500	6900				21900								
1	4700									9500					47000		
2				7300			9850		5550	16200				15800			
3								4500	37500	0			4300	22800			58500
4									10500	6700							
...
1996							19500	12350					11900				
1997						23999											
1998									14500				41500				
1999							16555					20500					
2000								4700		11000						6261.45	22800

Середнє.

```
[36]: pd.pivot_table(df1, values='price', index=['buyer'], columns=['year'], aggfunc=np.mean, fill_value='')
```

```
[36]:
```

	year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
buyer																	
0				7500	6900				21900								
1	4700									9500					47000		
2				7300			9850		5550	16200				15800			
3								4500	25350	0			4300	18500			58500
4									10500	6700							
...
1996							19500	12350					11900				
1997						13556.3											
1998									14500				41500				
1999							16555					18000					
2000								4700		11000						6261.45	22800

Кількість.

```
[37]: pd.pivot_table(df1, values='price', index=['buyer'], columns=['year'], aggfunc='count', fill_value='')
```

```
[37]:
```

year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
buyer																
0			1	1				1								
1	1								1					1		
2				1		1		1	1				1			
3							1	2	1			1	2			1
4								1	1							
...
1996						1	1					1				
1997					3											
1998								1				1				
1999						1					2					
2000							1		1						1	1

Вибрати топ-10 цін.

```
[40]: df.sort_values(by='price', ascending=False).head(10)
```

```
[40]:
```

	buyer	car	price	body	mileage	engV	engType	registration	year	model	drive
7621	1154	Bentley	547800.0	sedan	0	6.75	Petrol	yes	2016	Mulsanne	rear
7914	47	Bentley	499999.0	crossover	0	6.00	Petrol	yes	2016	Bentayga	full
1611	156	Bentley	499999.0	crossover	0	6.00	Petrol	yes	2016	Bentayga	full
4134	1438	Bentley	449999.0	crossover	0	6.00	Petrol	yes	2016	Bentayga	full
4325	767	Mercedes-Benz	300000.0	sedan	68	6.00	Petrol	yes	2011	S 600	NaN
5849	1182	Mercedes-Benz	300000.0	other	37	5.00	Petrol	yes	2012	G 500	full
1891	429	Mercedes-Benz	295000.0	sedan	29	6.00	Petrol	yes	2011	S 600	rear
2165	1946	Mercedes-Benz	295000.0	sedan	29	6.00	Petrol	yes	2011	S-Guard	rear
8205	1153	Land Rover	285000.0	crossover	0	5.00	Petrol	yes	2016	Range Rover	full
1478	1763	Bentley	259000.0	sedan	0	6.00	Petrol	yes	2014	Flying Spur	full

Вибрати топ-10 покупців за кількістю покупок.

```
[47]: df[['buyer', 'car']].groupby(['buyer']).count().sort_values(by='car', ascending=False).head(10)
```

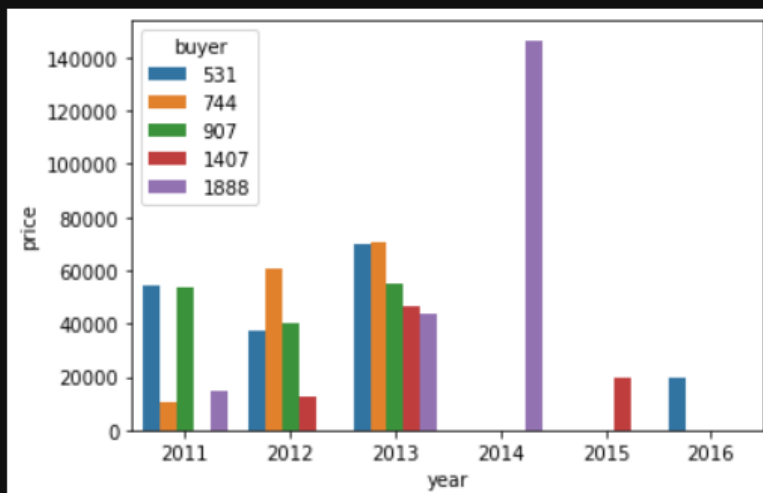
```
[47]:
```

	car
buyer	
412	15
1311	14
1979	13
907	13
355	13
1749	12
398	12
1641	12
920	12
1616	12

Побудувати крос-діаграму.

```
[136]: sns.barplot(x='year', y='price', hue='buyer', data=df3, estimator=sum, ci=None)
```

```
[136]: <AxesSubplot:xlabel='year', ylabel='price'>
```



Висновок

За час виконання даної лабораторної роботи я підготувала і завантажила дані пов'язані з продажем машин, вибирала дані за заданим критерієм, побудувала крос таблиці з різними агрегованими функціями та крос діаграму.