

Rapport du Projet BI : Business Intelligence

<< Analyse du marché de la location d'appartements au Maroc >>



Réalisé par :
Satir Abdellatif

A.U : 2022/2023

1- Idée de projet et problématique :

Récemment, après la période de quarantaine, le monde a été témoin de nombreux changements affectant toutes les régions. La chose la plus importante est peut-être la numérisation des choses, car tout est devenu distant (études, travail, ...) et le secteur du commerce et des affaires n'a pas été épargné par ces changements, car l'achat et la vente se sont propagés via Internet, et nous prenons par exemple le marché de la location d'appartements au Maroc, où le propriétaire peut la maison propose son service dans web et n'importe qui peut postuler à cette offre de location.

Mener une étude de marché sur les prix des appartements en location dans différentes villes est important pour plusieurs raisons :

- Il permet aux propriétaires et aux gestionnaires immobiliers de comprendre les conditions actuelles du marché dans différentes villes et d'ajuster leurs prix de location en conséquence pour rester compétitifs.
- Il permet aux locataires potentiels de prendre des décisions éclairées sur leur lieu de résidence en fonction du coût du loyer dans différentes villes, ce qui peut avoir un impact significatif sur leur budget global.
- Cela peut aider à identifier les zones d'une ville où les prix de location sont élevés ou bas, ce qui peut éclairer les futurs projets de développement ou de réaménagement.
- Cela peut également aider à prévoir les tendances futures des prix de location et permettre aux propriétaires et aux gestionnaires immobiliers de prendre des décisions plus éclairées concernant l'achat ou la vente de propriétés.
- Cela peut également aider le gouvernement et les décideurs politiques à prendre les bonnes décisions en matière de politique du logement, en étudiant les prix des appartements en location dans différentes villes, ils peuvent comprendre l'écart entre l'offre et la demande et la répartition de la richesse immobilière dans le pays.

- Il peut aider les investisseurs immobiliers à identifier les zones rentables dans lesquelles investir, en étudiant les prix de la location d'appartements dans différentes villes, ils peuvent identifier où les prix sont plus élevés et où ils peuvent générer plus de revenus.

Dans l'ensemble, les études de marché sur les prix de la location d'appartements dans différentes villes sont un outil important pour comprendre l'état actuel du marché du logement et prendre des décisions éclairées concernant la location, l'achat et l'investissement dans des propriétés.

2- Collection des données (Web Scraping):

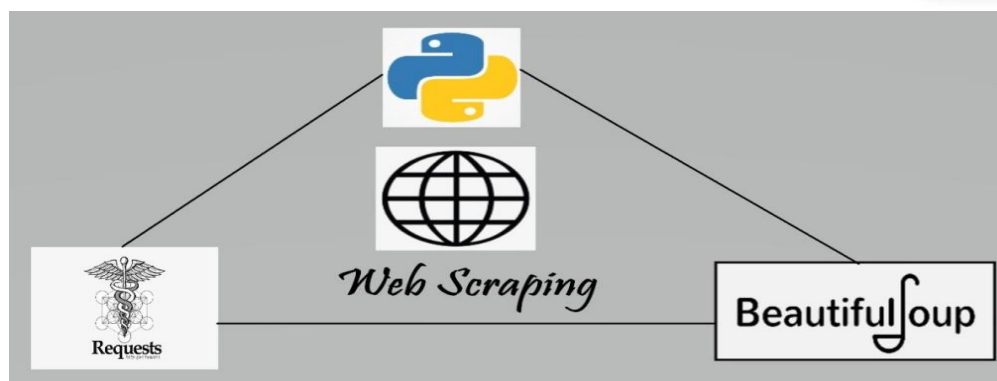
Qu'est-ce que le web scraping et pourquoi en ai-je besoin ?

La réponse simple est la suivante : tous les sites Web ne disposent pas d'une API pour récupérer le contenu. Vous voudrez peut-être obtenir des recettes de votre site Web de cuisine préféré ou des photos d'un blog de voyage. Sans API, l'extraction du HTML, ou le scraping, pourrait être le seul moyen d'obtenir ce contenu. Je vais vous montrer comment faire cela en Python.

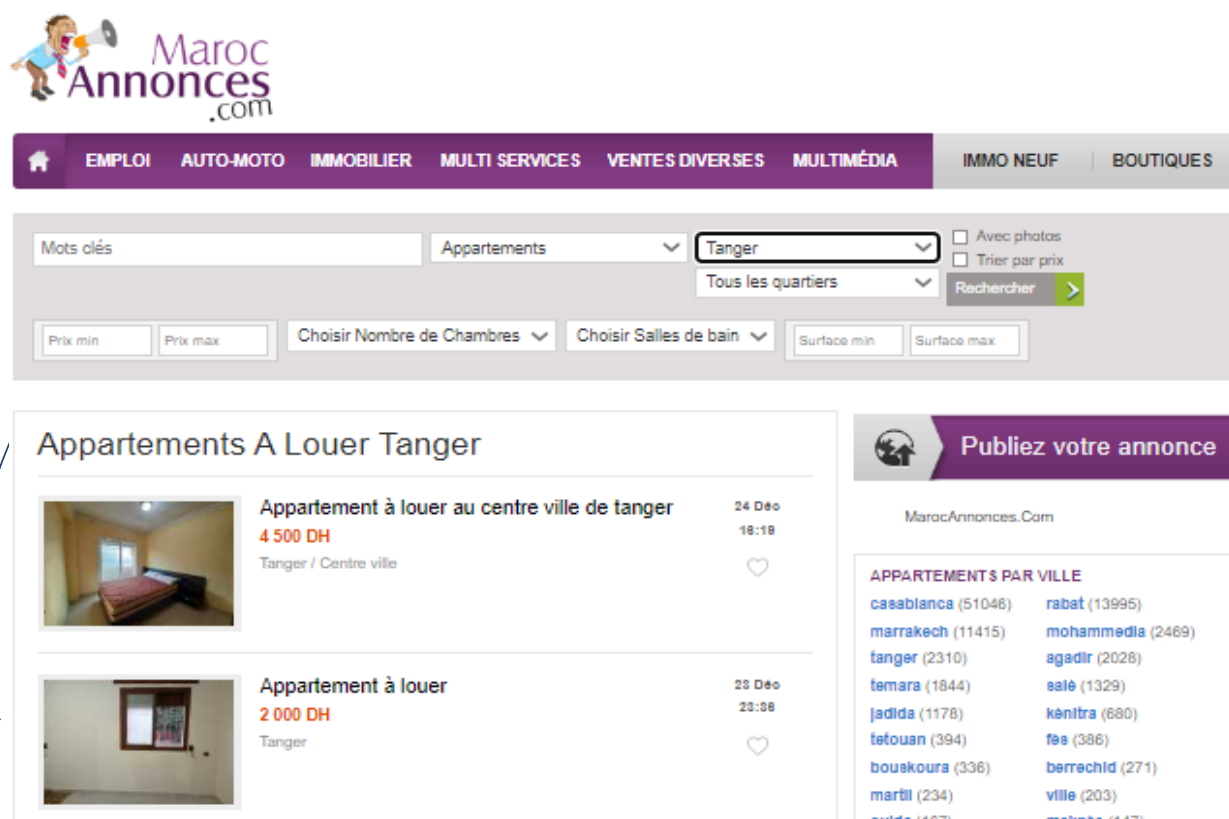
```
from bs4 import BeautifulSoup as bs
import requests as req
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
plt.style.use('ggplot')
```

Pour que le web scraping fonctionne en Python, nous allons effectuer les étapes de base :

- Extraction le contenu HTML à l'aide de la bibliothèque de Requests.
- Analysez la structure HTML et identifiez les balises qui ont notre contenu.
- Extraction les balises à l'aide de BeautifulSoup et placez les données dans une liste Python.



Dans ce projet on va extraire les données nécessaires dans le site web «marocannonces.com» qui offre plusieurs offres de location des appartements de différentes villes au Maroc



MarocAnnonces.com

EMPLOI AUTO-MOTO IMMOBILIER MULTI SERVICES VENTES DIVERSES MULTIMÉDIA IMMO NEUF BOUTIQUES

Mots clés Appartements Tanger Tous les quartiers Avec photos Trier par prix Rechercher

Prix min Prix max Choisir Nombre de Chambres Choisir Salles de bain Surface min Surface max

Appartements A Louer Tanger

Appartement à louer au centre ville de tanger 24 Déc 18:18
4 500 DH
Tanger / Centre ville

Appartement à louer 28 Déc 23:38
2 000 DH
Tanger

Publiez votre annonce

MarocAnnonces.Com

APPARTEMENTS PAR VILLE

casablanca (51046)	rabat (13995)
marrakech (11415)	mohammedia (2469)
tanger (2310)	agadir (2028)
temara (1844)	salé (1329)
jadida (1178)	kénitra (680)
tetouan (394)	fès (386)
bouekoura (336)	berrechid (271)
martil (234)	ville (203)
oujda (187)	meïna (147)

3- Collection et Chargement des données vers CSV files:

DATA SCRAPING

```
def cityy():

    print("Exp:AGADIR : https://www.marocannonces.com/maroc/location-appartements-agadir-b321-t552.html?pge=")
    url = input('URL : ')
    nb_page = int(input('Number of pages : '))
    filename = input('FileName : ')

    data = open('Rent_Analysis/'+filename+'.csv',mode="w",encoding="utf-8")
    header = "Title,City,Price,Time\n"
    data.write(header)

    for p in range(2,nb_page):
        link = req.get( f'{url}'+f'{p}' )
        sp = bs(link.content,'lxml')
        ul = sp.find('ul',class_="cars-list")
        li = ul.find_all('li')

        for i,info in enumerate(li):
            if len(info.get_text(strip=True)) != 0:
                info.extract()
                if (info.strong):
                    print(i,info.h3.text.strip(),',',info.span.text,',',info.strong.text,',',info.em.text.strip(),'\n')

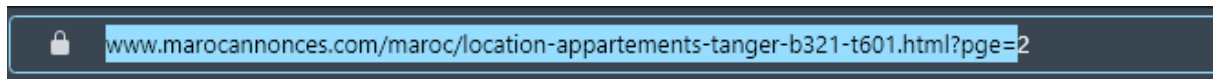
                    data.write(info.h3.text.strip() + ',' + info.span.text + ',' +
                               info.strong.text + ',' + info.em.text.strip() + '\n')

    data.close()

cityy()
```

Cette fonction demande à l'utilisateur de saisir :

- le lien du site Web (url) pour chaque ville.



Exemple : <https://www.marocannonces.com/maroc/location-appartements-tanger-b321-t601.html?pge=>

- le nombre de pages de cette ville qu'il voudra scrape (nb_pages).



- le nom du fichier qui sera stocké localement (filename).

Puis on crée un fichier csv (filename.csv) contenant dans son en-tête 4 colonnes : << Titre , Ville, Prix, Heure >>.

Et commence à extraire les données des pages entrées dans la fonction et stockez-les dans ce fichier csv.

Dans ce projet nous étudierons 6 grandes villes marocaines : Casablanca, Rabat, Tanger, Marrakech, Agadir et Kenitra.

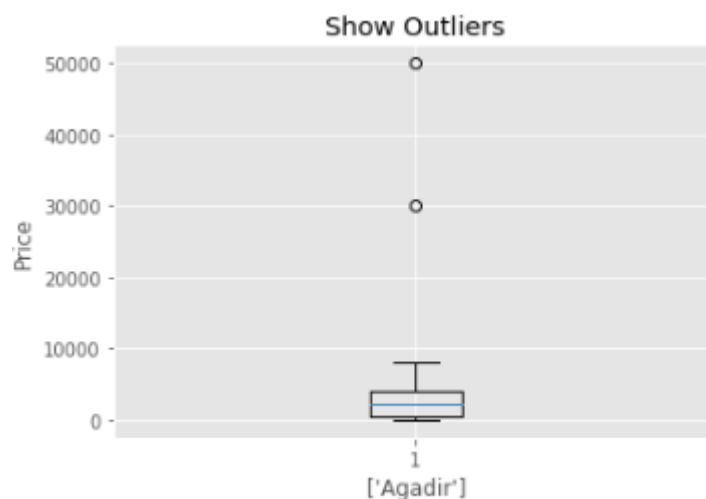
agadir	03/12/2022 16:20	Fichier CSV Micro...	8 KB
casa	03/12/2022 14:40	Fichier CSV Micro...	161 KB
cities_clean	30/10/2022 20:27	Fichier CSV Micro...	44 KB
kenitra	03/12/2022 16:30	Fichier CSV Micro...	6 KB
marrakech	03/12/2022 16:26	Fichier CSV Micro...	29 KB
rabat	03/12/2022 14:43	Fichier CSV Micro...	42 KB
tanger	03/12/2022 14:34	Fichier CSV Micro...	21 KB

4- Préparation et Nettoyage des données :

Après avoir collecté ces données, elles contiennent probablement beaucoup de valeurs aberrantes Outliers dans la colonne 'Prix' qui poseront un problème dans cette étude. Dans ce cas, nous devrions les détecter et les supprimer de nos données.

Il existe plusieurs façons de détecter et de supprimer les valeurs aberrantes dans les données :

- 1) **Inspection visuelle** : une façon de détecter les valeurs aberrantes consiste à inspecter visuellement les données à l'aide de graphiques tels que des histogrammes, des boîtes à moustaches ou des nuages de points. Les valeurs aberrantes apparaîtront généralement sous la forme de points de données significativement différents du reste des données.



- 2) **Z-score** : la méthode du Z-score calcule le nombre d'écarts types par rapport à la moyenne d'un point de données. Les points de données qui ont un Z-

score supérieur ou inférieur à un certain seuil (par exemple 3 ou -3) peuvent être considérés comme des valeurs aberrantes.

- 3) **Écart interquartile (IQR)** : La méthode IQR calcule l'écart entre le premier et le troisième quartile (Q1 et Q3) des données. Les points de données qui se situent en dehors de la plage de $Q1 - 1,5 * IQR$ à $Q3 + 1,5 * IQR$ peuvent être considérés comme des valeurs aberrantes.

DATA CLEANING ¶

```
def prepare(data):
    data['Price'] = data['Price'].astype('string')
    data['Price'] = data['Price'].str.replace('DH', '')
    data['Price'] = data['Price'].str.replace(' ', '')
    data['Price'] = pd.to_numeric(data['Price'])
    data['City1'] = data['City'].apply(lambda x: x.split('/')[0].replace(' ', ''))
    data['Street'] = data['City'].apply(lambda x: x.split(' / ')[-1])
    data.drop_duplicates(inplace=True)
    #show outliers
    plt.boxplot(data['Price'])
    plt.xlabel(data['City1'].unique())
    plt.ylabel('Price')
    plt.title('Show Outliers')
    plt.show()
    return data

# Method 2 : InterQuartile Range q3-q1
# sort -> calcul (q3-q1) -> find lower bound (q1*1.5) -> find upper bound (q3*1.5)
# -> Anything that lies outside of lower and upper bound is an outlier.
def inter_quartile(data):
    q1 = np.percentile(data['Price'], 25)
    q3 = np.percentile(data['Price'], 75)
    IQR = q3 - q1
    lower_bound = q1 - (1.5*IQR)
    upper_bound = q3 + (1.5*IQR)
    print('minor outliers : ', lower_bound, ', ', upper_bound, ']')
    #plot
    plt.xlabel(data['City1'].unique())
    plt.ylabel('Price')
    plt.title('Without Outliers')
    plt.boxplot(data.Price[data['Price'] <= upper_bound])
    plt.show()
    data = data[data['Price'] <= upper_bound]
    return data
```

```
def clean(data):
    print(data.shape)
    data = prepare(data)
    z_score(data)
    data = inter_quartile(data)
    print(data.shape)
    return data

casa = clean(casa)
casa
```

On a 3 trois fonctions :

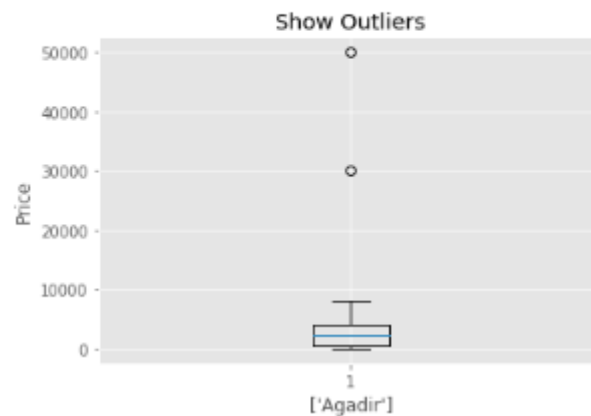
- la première fonction '*prepare*' est une fonction qui nous permet de préparer notre donnée par exemple dans la colonne Prix (supprimer DH et

convertir-la au numéro), et la division de la colonne city en deux parties (street nom, city nom et affichage de la visualisation de données à l'aide de box plot pour voir outliers.

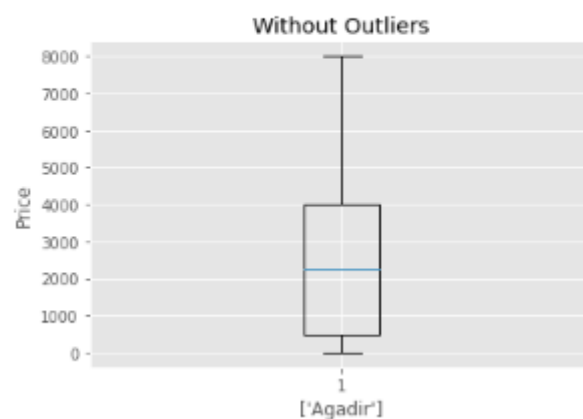
- la deuxième fonction '*inter_quartile*' qui nous permet de détecter et supprimer outliers dans notre données à l'aide de la méthode d'Écart interquartile (IQR)
- La troisième fonction '*clean*' qui combine les deux fonctions précédentes et nous renvoie la nouvelle données (clean data)

```
agadir = clean(agadir)
agadir
```

(98, 6)



```
outliers : [50000, 30000]
minor outliers : [ -4687.5 , 9212.5 ]
```



Save clean data

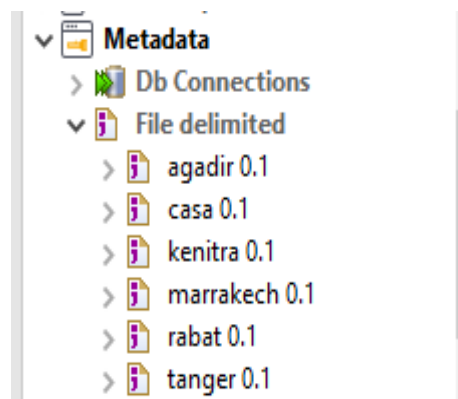
```
casa.to_csv('Rent_Analysis/Clean/casa.csv')
rabat.to_csv('Rent_Analysis/Clean/rabat.csv')
tanger.to_csv('Rent_Analysis/Clean/tanger.csv')
marrakech.to_csv('Rent_Analysis/Clean/marrakech.csv')
agadir.to_csv('Rent_Analysis/Clean/agadir.csv')
kenitra.to_csv('Rent_Analysis/Clean/kenitra.csv')
```


5- Transformation et Chargement dans ETL Talend et L'intégration dans Data Warehouse:

Talend. est un ETL (Extract Transform and Load) qui permet d'extraire des données d'une source, de modifier ces données, puis de les recharger vers un destination. La source et la destination des données peuvent être une base de données, un service web, un fichier csv...



Etape 1 : Création des métadonnées (chargement de notre clean data)



Etape 2 : Création la base de données

Server: 127.0.0.1

Databases SQL Status User accounts Export Import

Databases

Create database ?

Rent_dwh utf8mb4_general_ci Create

Etape 3 : Création des tables de dimension de chaque ville

Repository

LOCAL: Rent_Analysis2

- Job Designs
 - agadir 0.1
 - casa 0.1
 - Fact 0.1
 - kenitra 0.1
 - marrakech 0.1
 - rabat 0.1
 - tanger 0.1
- Contexts
- Code
- SQL Templates
- Metadata
 - Documentation
 - Recycle bin

Job casa 0.1 Job rabat 0.1 Job tanger 0.1 Job marrakech... Job agadir 0.1 Job kenitra 0.1

row1 (Main) tMap_1 dim_casa (Main order:1) show (Main order:2) tDBOutput_1 tLogRow_1

Schema editor Expression editor

row1

Column	Key	Type	N.	Date Pattern (...)	Length	Precision	Defa...	Comm...
id	<input checked="" type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		2	0		
Title	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		255	0		
City	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		29	0		
Price	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		4	0		

dim_casa

Column	Key	Type	N.	Date Pattern (...)	Length	Precision	Defa...	Comm...
id	<input checked="" type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		2	0		
Title	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		255	0		
Address	<input type="checkbox"/>	String	<input checked="" type="checkbox"/>		29	0		
Price	<input type="checkbox"/>	Integer	<input checked="" type="checkbox"/>		4	0		

Job(casa 0.1) Contexts(casa) Component x Run (Job casa)

tDBOutput_1(MySQL)

Basic settings

Database: MySQL

Property Type: Built-In

DB Version: Mysql 8

☐ Use an existing connection

Host: "localhost" * Port: "3306" *

Database: "rent_dwh" *

Username: "root" * Password: "*****" *

Table: "dim_casa" ...

Action on table: Drop table if exists and create Action on data: Insert

Schema: Built-In Edit schema Sync columns

Data source

This option only applies when deploying and running in the Talend Runtime

☐ Specify a data source alias

Etape 4 : Chargement les tables de dimension dans la base de données

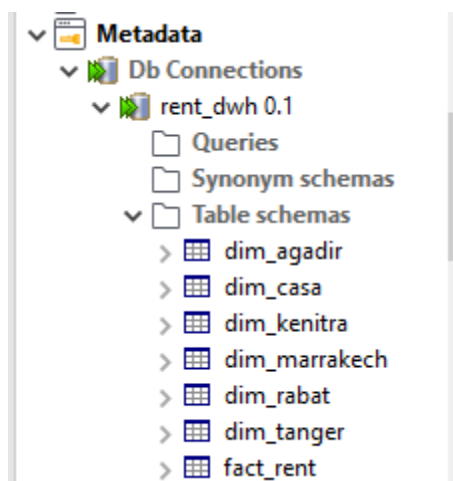


Table	Action	Rows	Type	Collation	Size	Overhead
<input type="checkbox"/> dim_agadir		96	InnoDB	utf8mb4_general_ci	16.0 KiB	-
<input type="checkbox"/> dim_casa		1,680	InnoDB	utf8mb4_general_ci	256.0 KiB	-
<input type="checkbox"/> dim_kenitra		71	InnoDB	utf8mb4_general_ci	16.0 KiB	-
<input type="checkbox"/> dim_marrakech		335	InnoDB	utf8mb4_general_ci	80.0 KiB	-
<input type="checkbox"/> dim_rabat		454	InnoDB	utf8mb4_general_ci	80.0 KiB	-
<input type="checkbox"/> dim_tanger		139	InnoDB	utf8mb4_general_ci	48.0 KiB	-
6 tables	Sum	2,775	InnoDB	utf8mb4_general_ci	496.0 KiB	0 B

1

>

>>

Number of rows:

25

Filter rows:

Search this table

Sort by key:

None

Options

				id	Title	Address	Price	Time	Street	City1
<input type="checkbox"/>				0	Appartement vide 150m² Florida Sidi Maarouf	Casablanca / Sidi Maarouf	6000	01 Déc 2022 13:23	Sidi Maarouf	Casablanca
<input type="checkbox"/>				1	Un appartement à louer pour un couple	Casablanca / Sidi Maarouf	4300	01 Déc 2022 10:38	Sidi Maarouf	Casablanca
<input type="checkbox"/>				2	CHERCHE APPART PRES DE FAC DE MEDECINE	Casablanca	4000	01 Déc 2022 10:38	Casablanca	Casablanca
<input type="checkbox"/>				4	Appartement à louer à Nour Californie	Casablanca / Ain Chok	6000	01 Déc 2022 10:37	Ain Chok	Casablanca
<input type="checkbox"/>				5	Un appartement à louer à lotissement hamza 100 m²	Casablanca / Sidi Maarouf	4000	01 Déc 2022 10:37	Sidi Maarouf	Casablanca
<input type="checkbox"/>				6	Part loue très bel appart ensoleillé bien agencé	Casablanca / Maarif	6300	01 Déc 2022 10:37	Maarif	Casablanca
<input type="checkbox"/>				7	Location	Casablanca / Errahma	1600	30 Nov 2022 17:30	Errahma	Casablanca
<input type="checkbox"/>				8	Appartement F3 NEUF 43m² 2ème étg. à Beauséjour	Casablanca / Beauséjour	5500	30 Nov 2022 13:35	Beauséjour	Casablanca
<input type="checkbox"/>				9	Location appartement mobilier	Casablanca	5000	30 Nov 2022 13:34	Casablanca	Casablanca
<input type="checkbox"/>				10	Appartement plein centre 2 Mars	Casablanca	5500	29 Nov 2022 23:33	Casablanca	Casablanca
<input type="checkbox"/>				11	Appart meuble bd med 5 4500 dh ttc	Casablanca / Centre ville	4500	29 Nov 2022 20:58	Centre ville	Casablanca
<input type="checkbox"/>				12	Appartement à louer à oulfa	Casablanca / Oulfa	3000	29 Nov 2022 20:57	Oulfa	Casablanca
<input type="checkbox"/>				13	BEL APPARTEMENT A DAR BOUAZZA	Casablanca / Dar Bouazza	8000	29 Nov 2022 20:32	Dar Bouazza	Casablanca
<input type="checkbox"/>				14	Appartement 90m² au 3ème ensoleillé Résd. Joumeira	Casablanca / Maarif	6000	29 Nov 2022 11:59	Maarif	Casablanca
<input type="checkbox"/>				15	Location appartement	Casablanca	1700	29 Nov 2022 11:59	Casablanca	Casablanca
<input type="checkbox"/>				16	Appartement a louer a oulfa	Casablanca / Oulfa	3500	28 Nov 2022 23:31	Oulfa	Casablanca
<input type="checkbox"/>				17	دائى منيدى كتيمة	Casablanca / Centre ville	1300	28 Nov 2022 19:44	Centre ville	Casablanca

6- Création des Dashboard Power BI :

Power BI est un puissant outil d'analyse et de visualisation de données qui peut être utilisé par un large éventail d'utilisateurs, des analystes commerciaux et des data scientistes aux responsables commerciaux et aux décideurs.

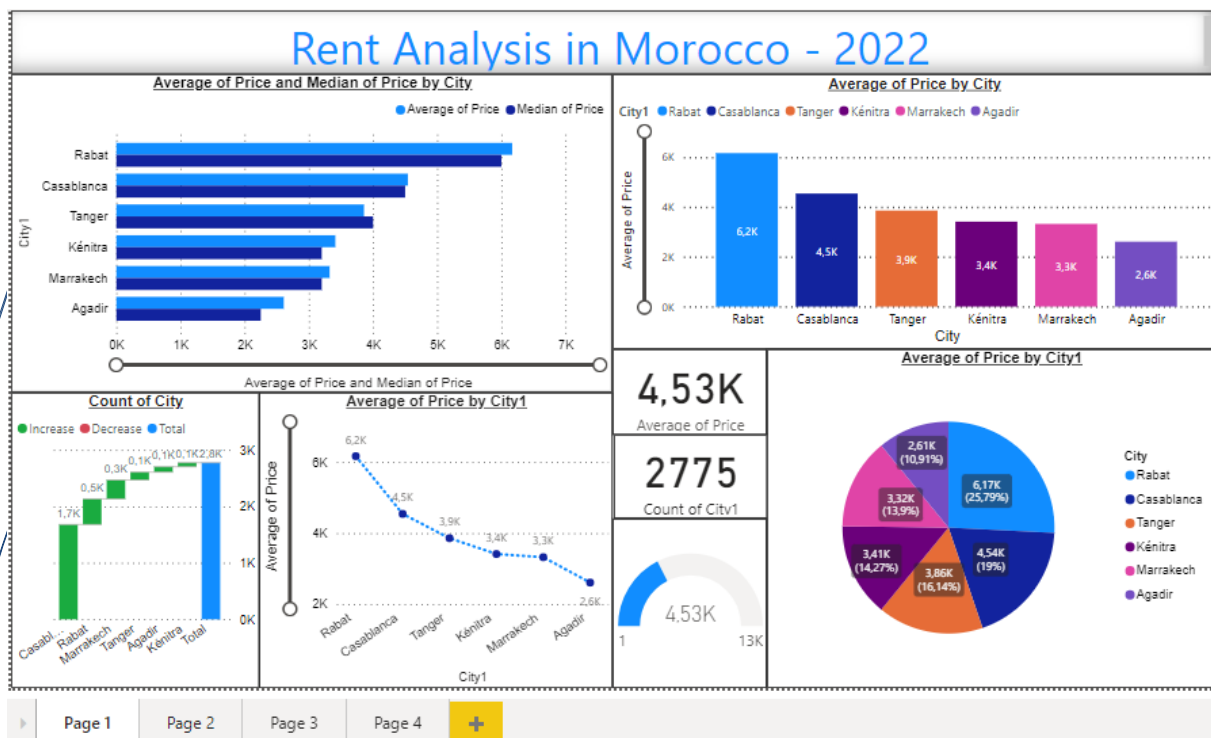


Power BI Desktop

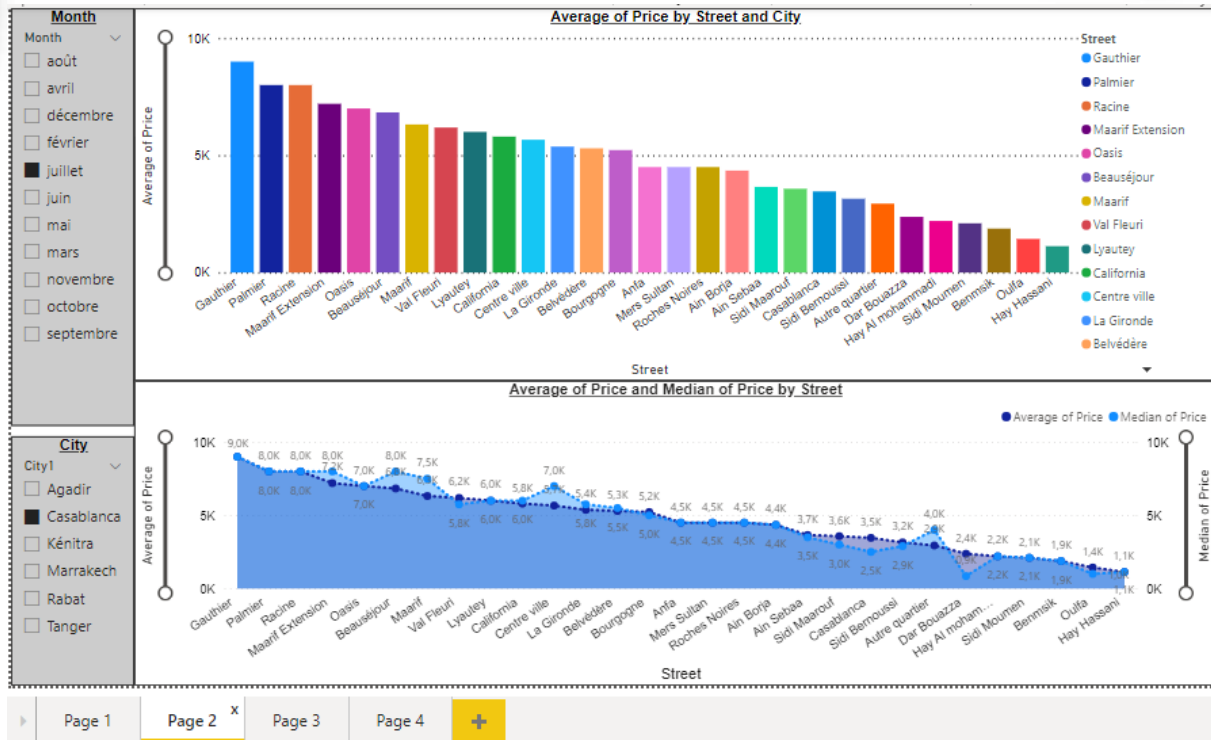
Etape 1 : Chargement des données et puis concaténation de toutes les tables de dimension et faire certaines transformations dans la colonne de Time

Title	Adress	Price	Street	City1	Date_Time	Date	Time	Month	Day	Hour
CHERCHE APPART PRES DE FAC DE MEDECINE	Casablanca	4000	Casablanca	Casablanca	01 Déc 2022 10:38	jeudi 1 décembre 2022	10:38:00	décembre	jeudi	10
Location appartement mobilier	Casablanca	5000	Casablanca	Casablanca	30 Nov 2022 13:34	mercredi 30 novembre 2022	13:34:00	novembre	mercredi	13
Appartement plein centre 2 Mars	Casablanca	5500	Casablanca	Casablanca	29 Nov 2022 23:33	mardi 29 novembre 2022	23:33:00	novembre	mardi	23
Location appartement	Casablanca	1700	Casablanca	Casablanca	29 Nov 2022 11:59	mardi 29 novembre 2022	11:59:00	novembre	mardi	11
Bel appartement ensoleillé 82m² à louer Belvedere	Casablanca	4800	Casablanca	Casablanca	28 Nov 2022 17:18	lundi 28 novembre 2022	17:18:00	novembre	lundi	17
Appartement Beau séjour	Casablanca	5000	Casablanca	Casablanca	26 Nov 2022 23:24	samedi 26 novembre 2022	23:24:00	novembre	samedi	23
Appartement haut. stand. 110 m2 la perle de californie	Casablanca	5800	Casablanca	Casablanca	23 Nov 2022 22:22	mercredi 23 novembre 2022	22:22:00	novembre	mercredi	22
Location long durée	Casablanca	2100	Casablanca	Casablanca	23 Nov 2022 14:18	mercredi 23 novembre 2022	14:18:00	novembre	mercredi	14
Location appartement	Casablanca	5000	Casablanca	Casablanca	22 Nov 2022 11:08	mardi 22 novembre 2022	11:08:00	novembre	mardi	11
Location long durée	Casablanca	2400	Casablanca	Casablanca	21 Nov 2022 20:16	lundi 21 novembre 2022	20:16:00	novembre	lundi	20
Appartte meuble neuf	Casablanca	4250	Casablanca	Casablanca	21 Nov 2022 15:41	lundi 21 novembre 2022	15:41:00	novembre	lundi	15
Appartement ensoleillé à 5min du tramway	Casablanca	2000	Casablanca	Casablanca	20 Nov 2022 20:45	dimanche 20 novembre 2022	20:45:00	novembre	dimanche	20
APPARTEMENT EL MENZEH	Casablanca	2800	Casablanca	Casablanca	19 Nov 2022 22:50	samedi 19 novembre 2022	22:50:00	novembre	samedi	22
شقة للكرام المهيبي	Casablanca	7000	Casablanca	Casablanca	19 Nov 2022 00:12	samedi 19 novembre 2022	00:12:00	novembre	samedi	0
شقة للكرام قرب محطة الطرامواي	Casablanca	2500	Casablanca	Casablanca	17 Nov 2022 04:59	jeudi 17 novembre 2022	04:59:00	novembre	jeudi	4
Des appartements de lux	Casablanca	500	Casablanca	Casablanca	15 Nov 2022 12:31	mardi 15 novembre 2022	12:31:00	novembre	mardi	12
Location d'un Appartement à Bernoussi	Casablanca	1800	Casablanca	Casablanca	14 Nov 2022 19:32	lundi 14 novembre 2022	19:32:00	novembre	lundi	19
Appartement bien entretenu	Casablanca	3500	Casablanca	Casablanca	14 Nov 2022 09:48	lundi 14 novembre 2022	09:48:00	novembre	lundi	9
Appartement propre ensoleillée à louer	Casablanca	1800	Casablanca	Casablanca	11 Nov 2022 22:52	vendredi 11 novembre 2022	22:52:00	novembre	vendredi	22
Appartement meublé à louer à Casablanca	Casablanca	400	Casablanca	Casablanca	11 Nov 2022 12:50	vendredi 11 novembre 2022	12:50:00	novembre	vendredi	12
Orthophoniste ou psychomotricienne	Casablanca	100	Casablanca	Casablanca	10 Nov 2022 12:45	jeudi 10 novembre 2022	12:45:00	novembre	jeudi	12
Appartement Meublé	Casablanca	6000	Casablanca	Casablanca	05 Nov 2022 08:47	samedi 5 novembre 2022	08:47:00	novembre	samedi	8
Appartement luxueux à louer sur Nouasseur	Casablanca	5500	Casablanca	Casablanca	04 Nov 2022 11:09	vendredi 4 novembre 2022	11:09:00	novembre	vendredi	11

Etape 2 : Les Dashboard

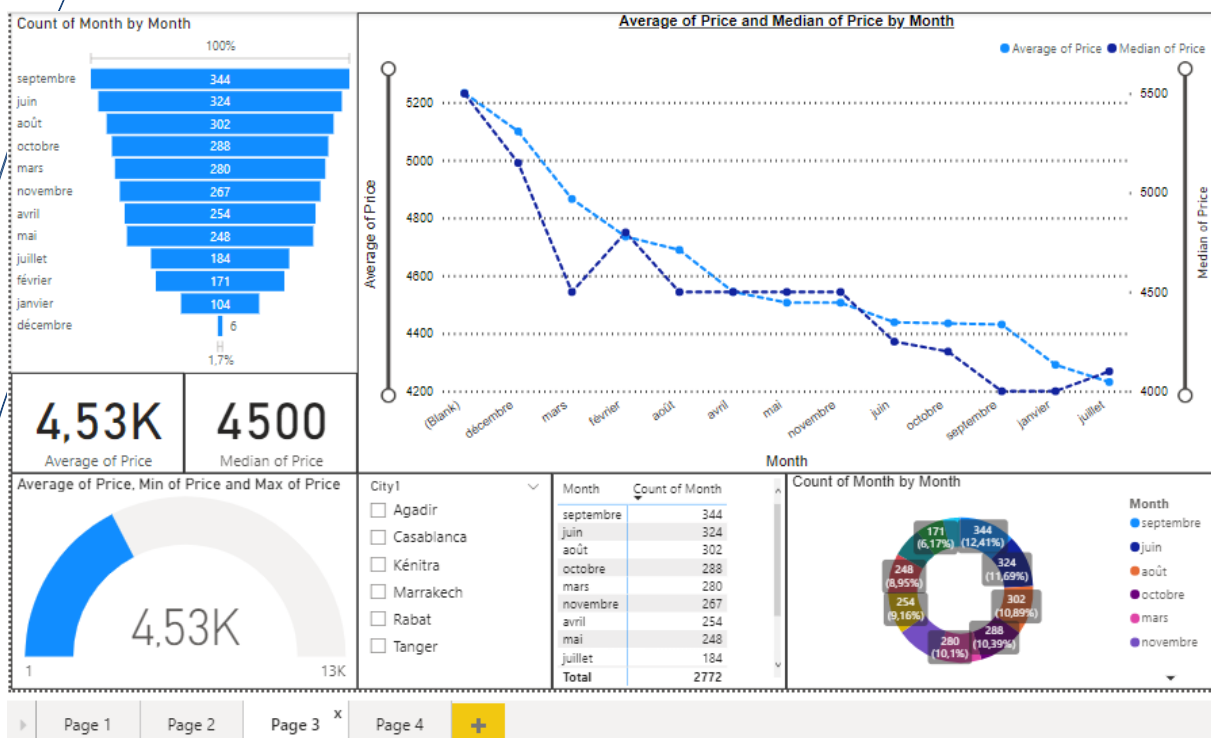


Comparaison de prix moyen de location des appartements
dans chaque ville.



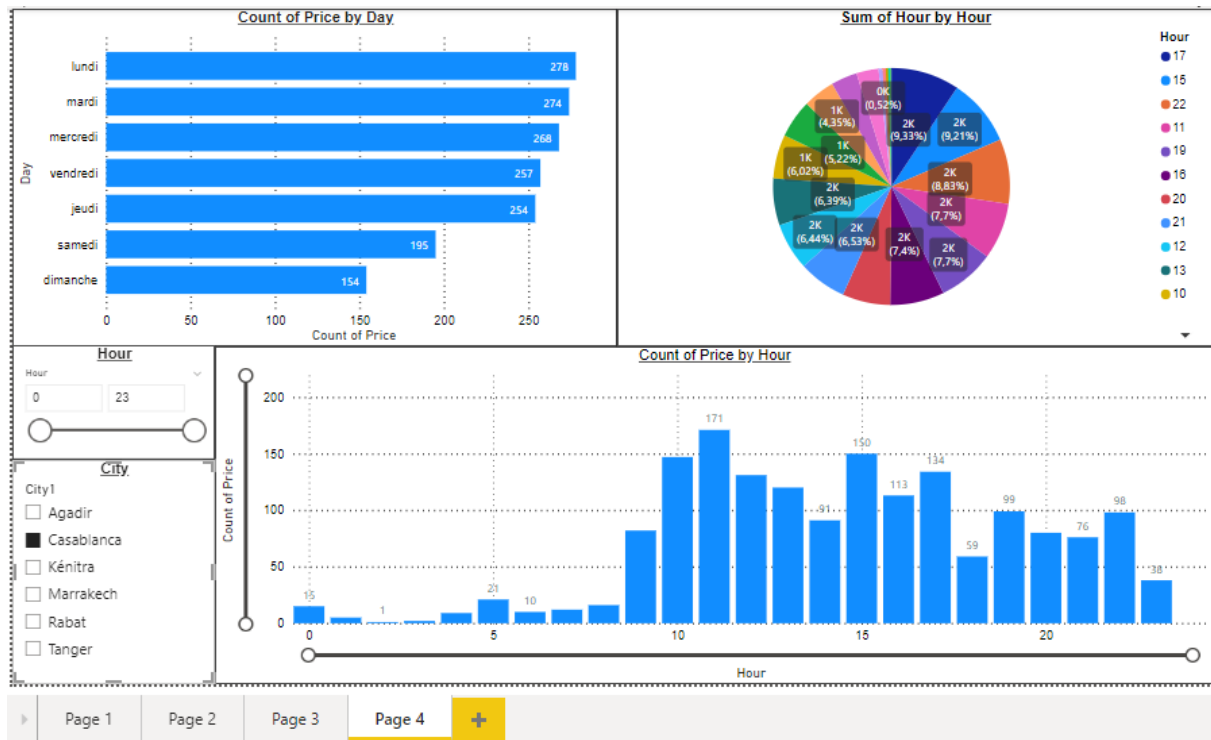
Comparaison de prix moyen de location des appartements

dans chaque quartier d'une ville sélectionné et chaque période sélectionné aussi.



Compter le nombre de données disponible de chaque mois

pour analyser le nombre des demandes ,et comparaison avec prix moyen de chaque mois.



Compter le nombre de données disponible de chaque jour,
et de chaque heure.

7- Conclusion:

Dans l'ensemble, les études de marché sur les prix de la location d'appartements dans différentes villes sont un outil important pour comprendre l'état actuel du marché du logement et prendre des décisions éclairées concernant la location, l'achat et l'investissement dans des propriétés...