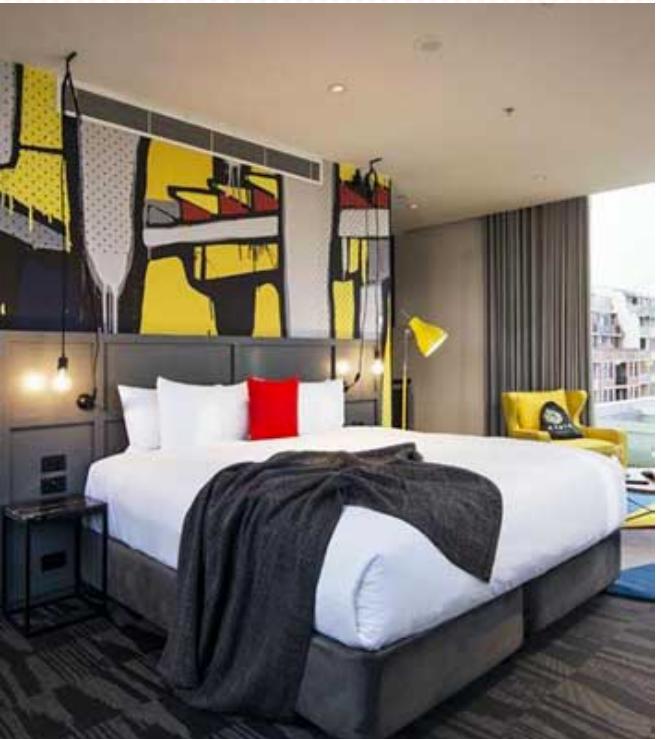


# SCRAPING PROJECT

ARTSPER x FOUR SEASON

ABELIA PETELLE – JANUARY 2020



- **CLIENT** : Four Seasons Hotels & Resort
- **CONTEXT** : For the opening of a new hotel in France, Four Seasons Hotels & Resorts wants to create a new concept and to have for each of their rooms a new theme regarding a chosen decade.
- **CLIENT'S EXPECTATION** : To know which artist to look at to fill their rooms with masterpieces that correspond to the chosen decade but also which are still trendy today.

Les garanties Artsper : paiements sécurisés, livraison dans le monde entier et retours gratuits ! [En savoir plus](#)

Accueil > Oeuvres > Edition > Downward Trajectory



Mise en situation

## Shepard Fairey (Obey)

*Downward Trajectory, 2019*

Edition

24 x 24 cm

750 €

Livraison : Une à deux semaines

+78 € de frais de livraison en France

Essayez gratuitement l'œuvre pendant 14 jours ⓘ

Ajouter au panier

♡ Ajouter à ma wishlist

1 exemplaire restant

5 personnes visitent cette page

# STRATEGY

Collect data from the website Artsper (marketplace who sells masterpieces of art from various wellknown galleries) in order to see which artists where/are the most present in these galleries.

# PROCESS FOLLOWED



**WEB SCRAPING**  
Artsper.com's URL & masterpieces pages



**DATA FRAME CREATION**  
with pandas



**EXPORT TO A DATABASE  
TO STOCK THE DATA**  
(MySQL)



**STANDARD CLEANING  
PROCESS**

Strip dates from names  
Convert French to English  
Create a decade column  
Order columns



**ANALYSIS**  
Filter the data regarding the  
decade & category of art  
input



**REPORTING**

# TOOLS USED

- **Libraries** : requests, BeautifulSoup, pandas, pymysql, create\_engine, os, matplotlib.pyplot, seaborn
- **Pipelines**
- **Functions** : lambda, apply, regex,
- **List comprehensions**
- **String and interger operations**
- **Error handling**

NAME	TYPE	DESCRIPTION
Artist	String	Name of the artist
Name	String	Name of the masterpiece
Date	Integer	Date of the edition of the masterpiece
Decade	Integer	Decade of the masterpiece (1900, 1910 .... 2010)
Category	String	Category of the masterpiece (Edition, Painting, Sculpture, Drawing, Photography)
Theme	Object	Theme of the masterpiece (Street art, Architecture, Paysage urbain, Nus)
Price	Object	Price of the masterpiece
Dimension	Object	Dimension of the masterpiece (HxLxP)
City Galery	Object	City and Country of the Galery
Name & Gallery	Object	Name and Date of the masterpiece

*Used for  
the  
analysis*

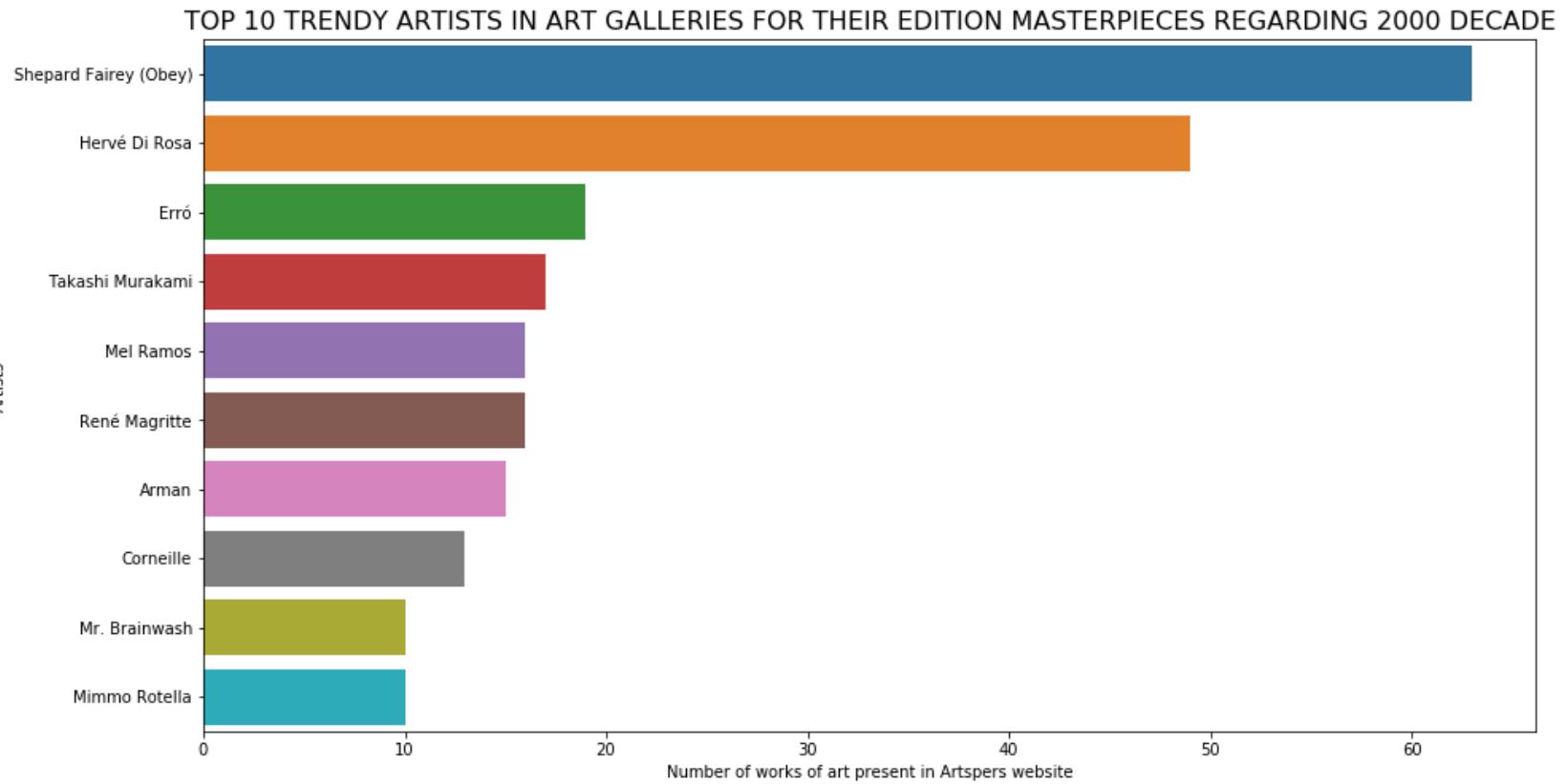
*Can be  
used for  
further  
analysis*

## FINAL DATAFRAME

4423 rows × 7 columns

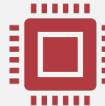
# OUTPUT / RESULT

*To fill the room  
« 2000's » with edition  
masterpieces, our  
recommandation is to  
be interested at the  
following artists*



Name	Type	Size	Value
category	str	1	Edition
data	DataFrame	(4423, 10)	Column names: Artist, Name & Date, Category, Price, Theme, Dimension, ...
decade	int	1	2000
df_artsper	DataFrame	(4423, 10)	Column names: Artist, Name, Date, Decade, Category, Theme, Price, Dime ...
first_filtered	DataFrame	(560, 10)	Column names: Artist, Name, Date, Decade, Category, Theme, Price, Dime ...
list_category	list	5	['Edition', 'Painting', 'Sculpture', 'Drawing', 'Photography']
list_decade	list	12	['1900', '1910', '1920', '1930', '1940', '1950', '1960', '1970', '1980' ...]
results	DataFrame	(10, 2)	Column names: Artists, Number of works of art present in Artspers webs ...
second_filtered	DataFrame	(342, 10)	Column names: Artist, Name, Date, Decade, Category, Theme, Price, Dime ...

# DELIVERABLES



## 2 Python (.py) code file :

- Artsper scraping pipeline
- Artsper data analysis pipeline



A **data folder** containing the data set.



An **output file** containing the histogram analysis (.png) regarding the decennie and category chosen.



A **README.md file** containing a detailed explanation of the process followed.

ANY  
QUESTIONS ?

