

e0

A series of thin, black, intersecting lines of various lengths and orientations, creating a complex, abstract geometric pattern that resembles a network or a series of overlapping planes. The lines are scattered across the upper and middle portions of the slide, with some lines extending towards the right edge.

# ANALISI DEI DATI PER LA PREVISIONE DELLE TENDENZE NEL SETTORE DEL FASHION

Riccardo Medici

## Diapositiva 1

---

**e0** Manca tutto il workflow della parte di estrazione dei dati; puoi anche mettere degli screen del sito e parlarci sopra spiegando come opera il crawler, ma non inserirla proprio è sbagliato

Mancano le conclusioni e sviluppi futuri

emanuele.balloni@unimc.it; 2022-09-27T16:03:47.844



# ANALISI DEI DATI PER LA PREVISIONE DELLE TENDENZE NEL SETTORE DEL FASHION

DATA ANALYSIS FOR FORECASTING TRENDS IN THE  
FASHION INDUSTRY.

Relatore: Emanuele Frontoni

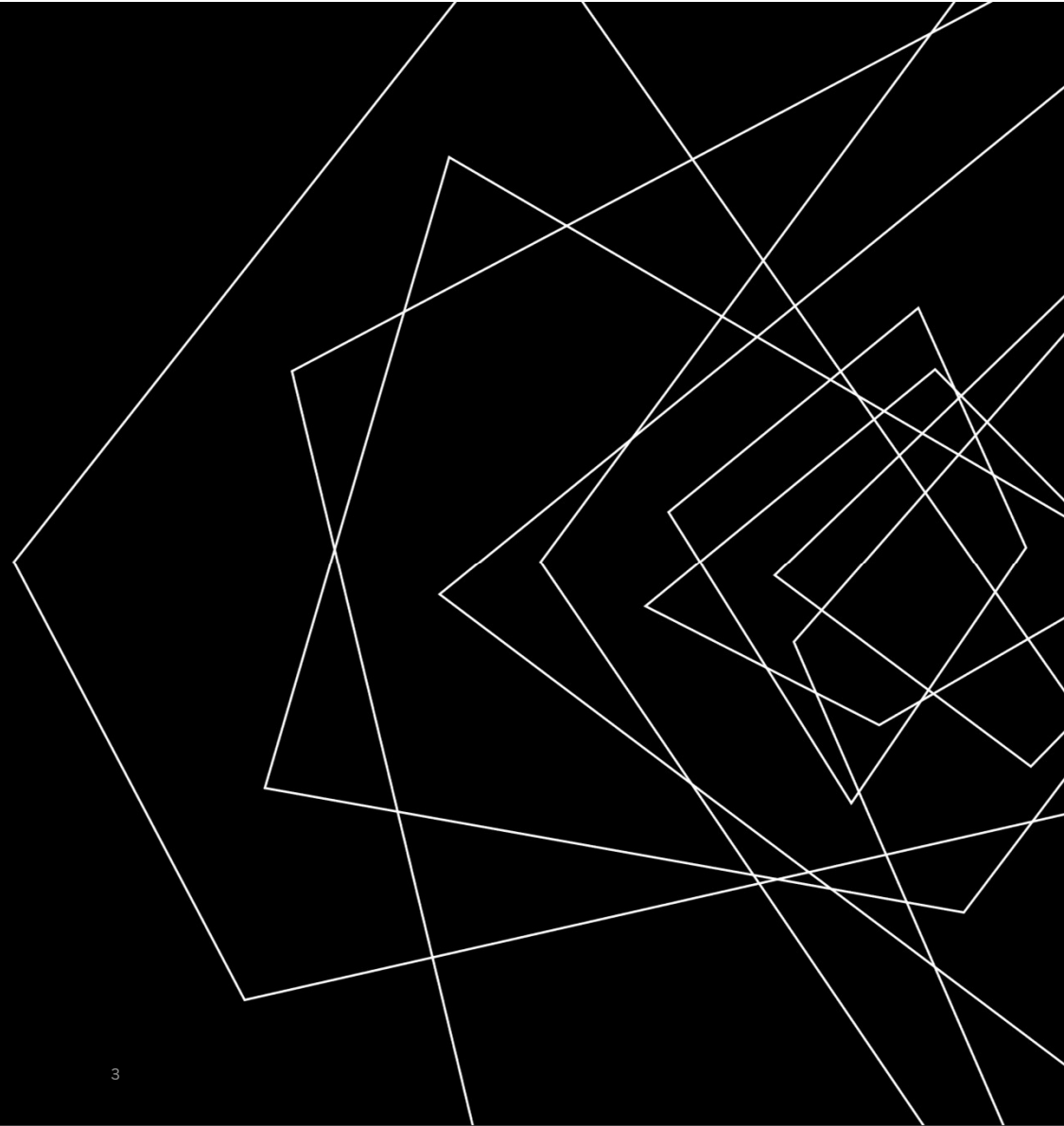
Correlatore: Marina Paolanti

Candidato:

Riccardo Medici

# SOMMARIO

Ogni settore che utilizzava dati storici e tecniche statistiche per tentare di prevedere il futuro prossimo ora ha accesso ad un pool di dati immenso. Questa nuova branca dello studio dei dati si chiama Data Science e per accedere ai dati presenti su internet esistono tecniche e metodologie specifiche. Un esempio di queste tecniche, usato in questo tirocinio, è il cosiddetto web Scraper, ovvero un programma utilizzato per raccogliere informazioni dai siti internet. Questo programma, visualizza gli elementi presenti su Farfetch e li memorizza in un file Json. Questi dati vengono poi estrapolati e analizzati attraverso dei grafici per dare uno spaccato della distribuzione di brand e prodotti sull' e-commerce.



### Diapositiva 3

---

- e0**      **analizzati attraverso**  
emanuele.balloni@unimc.it; 2022-09-27T15:50:08.905
- e1**       **tirocinio**  
emanuele.balloni@unimc.it; 2022-09-27T15:52:51.577

INTRODUZIONE

CONTESTO, OBIETTIVI DELLA TESI

e0

STATO DELL'ARTE

IMPORTANZA DATA ANALISI, WEB SCRAPER, CONFRONTO METODI

MATERIALI E  
METODI

PYTHON, LIBRERIE UTILIZZATE, FUNZIONI UTILIZZATE,  
PROBLEMI RISCONTRATI

TABELLE E RISULTATI

GRAFICI GENERATI DAI DATI E CONCLUSIONI

## Diapositiva 4

---

e0

cancella

emanuele.balloni@unimc.it; 2022-09-27T15:53:10.328

# INTRODUZIONE

## CONTESTO

- Essere in grado di prevedere l'andamento del mercato offre un grande vantaggio rispetto alla concorrenza
- Tecniche previsionali hanno avuto una grande evoluzione ma si basavano sempre e solo su dati storici
- La Data Science è una nuova branca dello studio di dati che utilizza informazioni di diverso tipo non solo dati storici

## OBBIETTIVI

- Estrapolare dati di borse dall'e-tailer Farfetch
- Costruire un file Json contenente dati relativi ad ogni borsa analizzata
- Generare grafici per comprendere meglio l'analisi di mercato svolta

e0



## Diapositiva 5

---

**e0**

Analisi della

emanuele.balloni@unimc.it; 2022-09-27T15:50:56.607

**e1**

Togliere tutta questa parte, questa non è la tesi ma la presentazione del lavoro svolto nel tirocinio

emanuele.balloni@unimc.it; 2022-09-27T15:51:39.732



# STATO DELL'ARTE

## IMPORTANZA DATA ANALYSIS

e0

- Vantaggi e ritorni per le aziende
- Ostacoli all'adozione
- Business intelligence
- Applicazioni

## WEB SCRAPING

- Come funziona un web scraper
- Le diverse tecniche di webscraping

## CONFRONTO REGEX, XPATH E HTML DOM

- Cosa sono xpath, html dom e regex
- Paper dell' 'Atlantis press' in cui vengono messe a confronto le performance dei tre metodi.

## Diapositiva 6

---

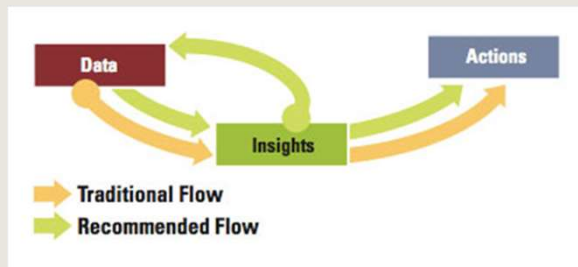
e0

O scrivi Data analysis o Analisi dei dati

emanuele.balloni@unimc.it; 2022-09-27T15:54:07.187

# IMPORTANZA DATA ANALISI

- L' utilizzo della data science permette di ottenere delle informazioni sugli del proprio settore molto più precisi rispetto all'utilizzo dei soli dati storici.
- Secondo uno studio del MIT le aziende più performanti usano massicciamente la data science.
- Alcuni esempi delle applicazioni della data analisi: Marketing, Finanza e Manifattura.



# WEB SCRAPING

- Si definisce web scraping una tecnica per leggere e estrarre informazioni da un sito web.
- Ci sono diverse tecniche di web scraping: Metodo mimico, Weight Measurement Approach, Metodo differenziale, Machine Learning.



# CONFRONTO REGEX, XPATH E HTML DOM

- Un paper di ricerca pubblicato su 'Atlantis Press' ha messo a confronto tre metodi utilizzati per accedere agli elementi di una pagina html
- **Xpath:** Xpath è un indirizzo di navigazione univoco per ogni elemento
- **HTML DOM:** è uno standard per accedere, modificare e cancellare gli elementi presenti in una pagina html.
- **Regex:** metodo per identificare un elemento tramite un determinato gruppo di parole.
- Dai risultati dei test Regex ha ottenuto un tempo medio di risposta di **399.75ms**, the DOM HTML method ha una media di **298.55ms** e il metodo Xpath ha una media di **435.15ms**

Experiment	Time (ms)		
	REGEX	HTML DOM	XPATH
1	375	297	406
2	375	250	407
3	390	360	485
4	391	281	859
5	391	454	422
6	382	265	390
7	446	266	406
8	322	281	422
9	325	265	406
10	377	500	438
11	294	250	391
12	286	265	375
13	422	266	640
14	390	266	390
15	516	250	375
16	406	266	375
17	406	250	375
18	594	282	375
19	391	407	391
20	516	250	375
Avg	399,75	298,55	435,15



# MATERIALI E METODI

e1

## PYTHON

Python è un linguaggio di tipo dinamico, utilizzato nel progetto per la sua versatilità e facilità d'uso.

## LIBRERIE DI PYTHON UTILIZZATE

Selenium, Pandas, Time e datetime, Json, WebDriver

## DEFINIZIONI UTILI e FUNZIONI WEBDRIVER

Raccolta di vocaboli tecnici utilizzati all'interno della tesi.

Funzioni usate all'interno del codice appartenenti alla libreria web driver.

## SFIDE INCONTRATE DURANTE LO SVILUPPO DEL CODICE

e0

Insieme di problematiche affrontate durante la scrittura del codice.

## Diapositiva 10

---

- e0**      **Sfide incontrate e superamento**  
emanuele.balloni@unimc.it; 2022-09-27T15:55:46.308
- e1**      **Materiali e metodi**  
emanuele.balloni@unimc.it; 2022-09-27T15:57:13.032



# LIBRERIE DI PYTHON UTILIZZATE

- **Selenium:** Selenium è una libreria Python open source che permette agli utenti di accedere ad una pagina web utilizzando l'interfaccia web driver. In questo modo è possibile navigare su una pagina html utilizzando funzioni che emulano il comportamento umano.
- **Pandas:** Pandas è una libreria Python open source utilizzata per la manipolazione dei dati e la loro analisi. Permette di creare diverse tipologie di grafici e tabelle in modo molto più rapido rispetto a pratiche più tradizionali come Excel.
- **Json:** Json è un formato per la scrittura e trasporto dei dati.



# FUNZIONI WEBDRIVER

•**driver.get()**: la funzione `.get()` prende come argomento un indirizzo url, che viene caricato su una pagina del browser quando essa viene chiamata.

•**WebDriverWait**: questa funzione aspetta s secondi finché l'elemento contrassegnato dal Xpath corrispondente risulta cliccabile o è individuata.

•**driver.find\_element()**: questa funzione trova l'elemento corrispondente all'argomento della funzione

•**driver.execute\_script()** : `execute_script()` esegue comandi JavaScript nella pagina utilizzata, nel caso specifico «`window.scrollTo`»

•**webdriver.ActionChains()**: le actionchains sono un metodo per automatizzare interazioni di "basso livello"

•**driver.sendkeys()**: questa funzione permette di inviare un input di tasti

•**driver.click**: si esegue una pressione del tasto sinistro del mouse.



## WORKFLOW CODICE

1. Selezione tipologia di borsa
2. Informazione sul numero di pagina e selezione modello
3. Visualizzazione singolo modello e acquisizione dati

## SELEZIONE TIPOLOGIA BORSA

**Donna** Uomo Bambino Beauty

**FARFETCH**

Novità Acquista per Brand Abbigliamento Scarpe Borse Accessori Gioielli Pre-owned Saldi

### BORSE

Tutte le borse

Borse da mare

Borse a secchiello

Clutch

Borse mini

Borse a spalla

Borse tote

Borse Pre-owned

Esclusive

Borse Conscious

Borse a tracolla

### COLLEZIONI ICONICHE

Tutte le borse iconiche

Balenciaga Le Cagole

Balenciaga Hourglass

Chanel 2.55 pre-owned

Gucci GG Marmont

Hermès Birkin pre-owned

Jacquemus Le Chiquito

Prada Cleo

Saint Laurent LouLou

Valentino Rockstud

Versace La Medusa

### VENDI E RIPARA

Vendi le tue borse

Ripara e restaura i tuoi  
preferiti

Positively FARFETCH

# SELEZIONE MODELLO BORSA

## BORSE A SECCHIELLO

Le borse a secchiello da donna sono l'accessorio su cui investire per la nuova stagione. In pelle color pastello o in paglia, morbida, strutturata, mini o XXL, abbinala a tutti i tuoi look. Da [Gucci](#) ad [Alexander](#). [Mostra di più](#)

Filtri <

Ordina per ▾

Brand ▾

Colori ▾

Prezzo ▾

Consegna In Giornata ▾

Conscious ▾



Nuova stagione  
**Prada**  
Borsa a secchiello

**Prada**  
Borsa con coulisse

Nuova stagione  
**ETRO**  
Borsa a secchiello con stampa

# ACQUISIZIONE DATI DEL SINGOLO MODELLO



Nuova stagione  
**Prada**  
Borsa a secchiello

1.300 €  
IVA inclusa

Taglia unica disponibile

**Aggiungi al carrello**

CONSEGNA PREVISTA  
3 ott - 6 ott

DISPONIBILE ANCHE IN



## DETTAGLI PRODOTTO

## TAGLIA E VESTIBILITÀ

## SPEDIZIONE E RESI

Nuova stagione  
**Prada**  
Borsa a secchiello

Nero, rifiniture in pelle, triangolo in smalto con logo, dettagli color oro, manico singolo, chiusura a coulisse, scomparto frontale con zip e tracolla regolabile e rimovibile. POSITIVELY CONSCIOUS: questo prodotto contiene ECONYL® – un tessuto di nylon ricavato da materiali rigenerati come reti da pesca, tappeti e altri rifiuti che possono essere riciclati e ricreati – che contribuisce a ridurre l'impronta ecologica.

## Composizione

Esterno: Tessuto 100%

**ID brand:** 1BH038VUOORV44



## OSTACOLI SUPERATI DURANTE LO SVILUPPO DEL CODICE

- Cookies and ads
- Utilità del try except
- Utilità del webdriver wait
- Cambio pagina
- Selezione tipologia borsa

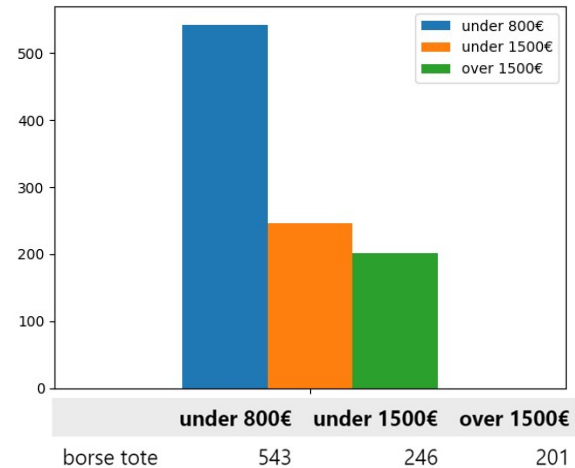
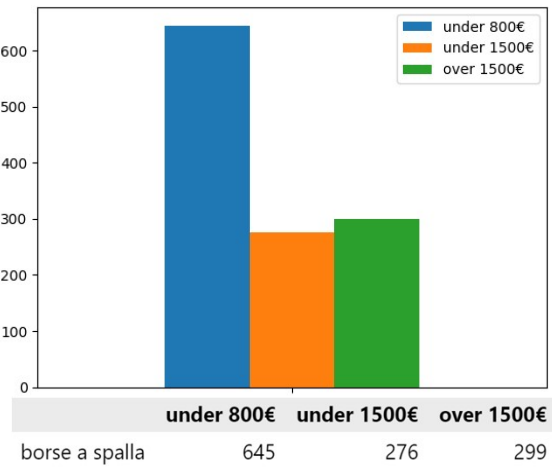
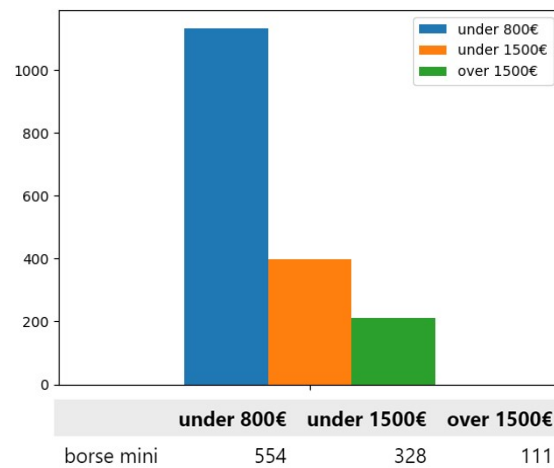
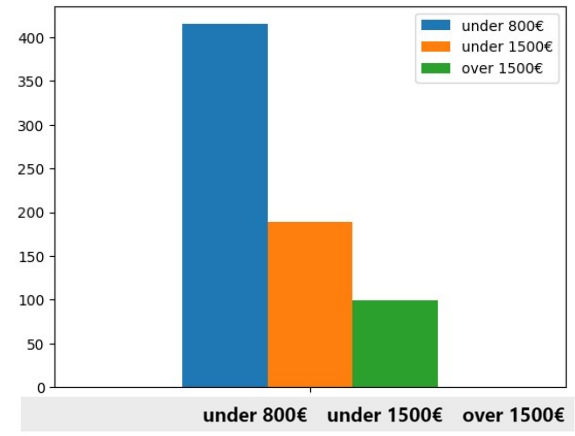
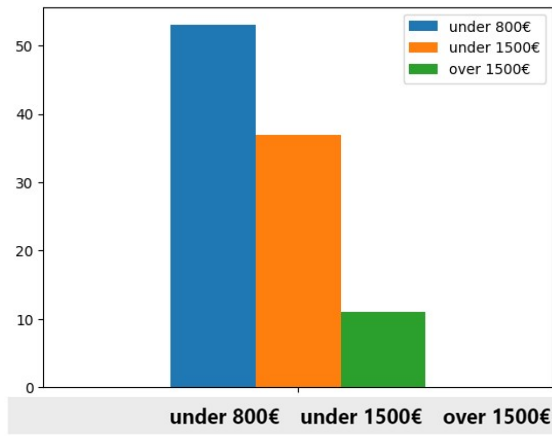
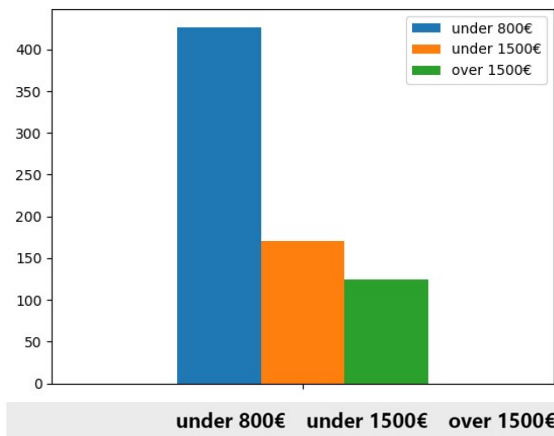


## TABELLE E RISULTATI

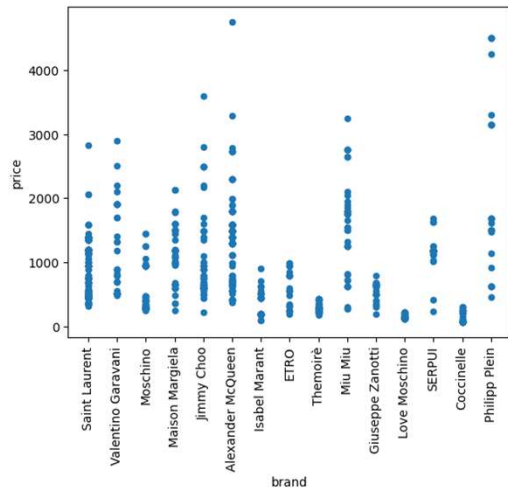
Dai dati presenti nel file Json sono stati generate dei grafici; i quali confrontano diversi parametri per evidenziare determinate informazioni

- **Numero di borse per fascia di prezzo per categoria**  
Istogramma, fasce sotto 800€, sotto 1500€, sopra 1500€
- **Range di prezzo per brand per ogni categoria**  
Grafici a dispersione
- **Numero di borse per marchio per fascia di prezzo**  
Istogrammi orizzontali decrescenti
- **Prezzo medio per brand per categoria**  
Istogrammi, rappresentati i 15 brand più diffusi per grafico

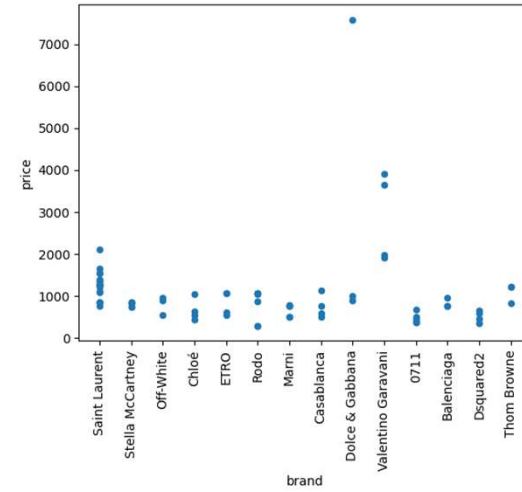




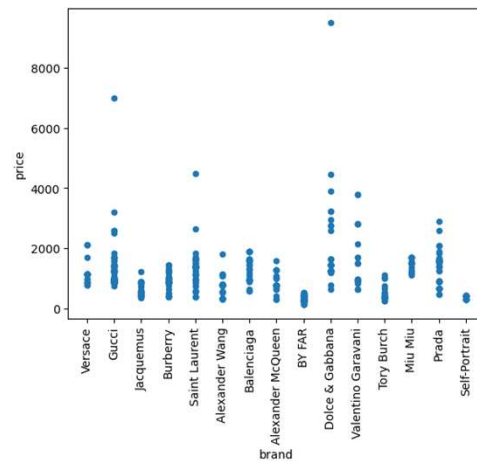
Confronto Prezzo/Brand - Borse Clutch



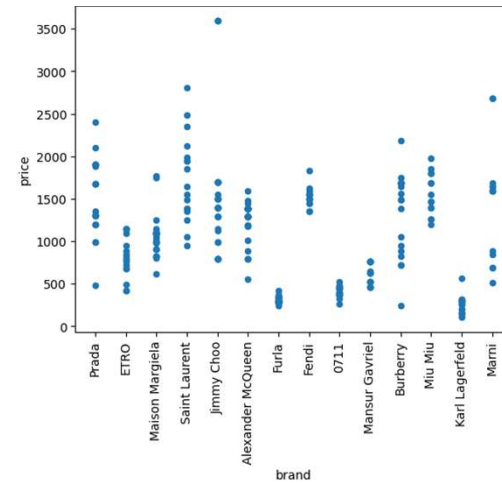
Confronto Prezzo/Brand - Borse da mare

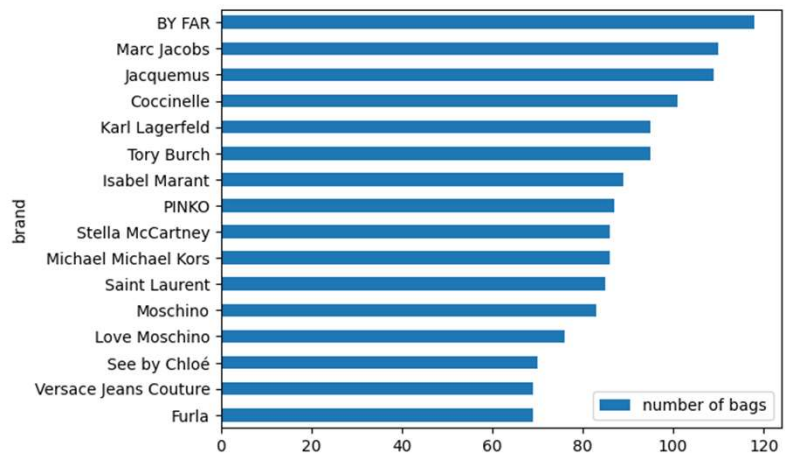


Confronto Prezzo/Brand - Borse mini

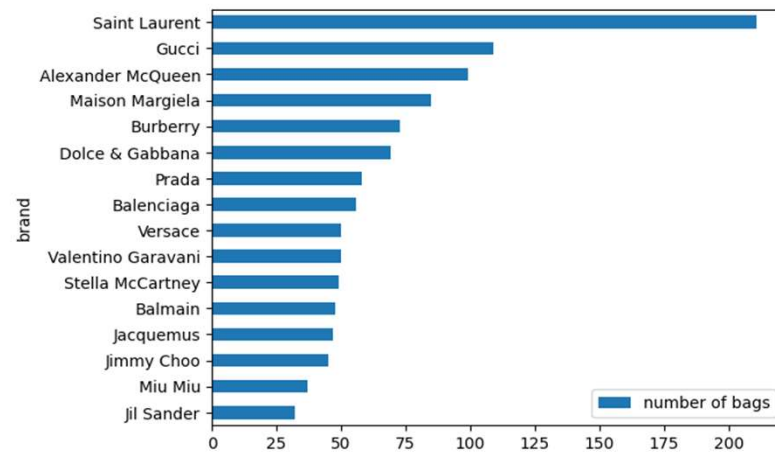


Confronto Prezzo/Brand - Borse a secchiello

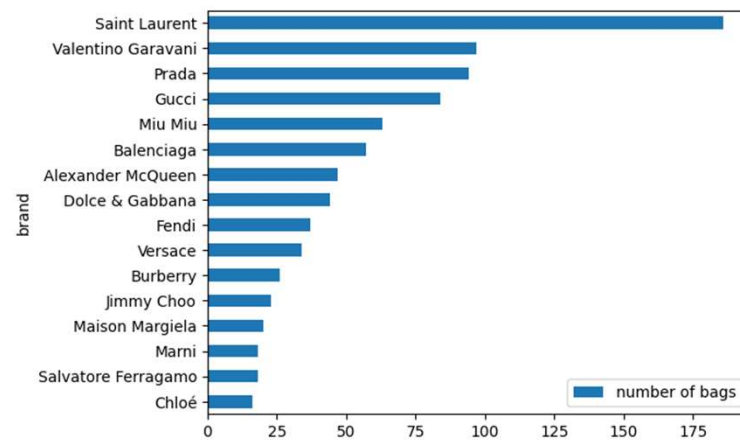




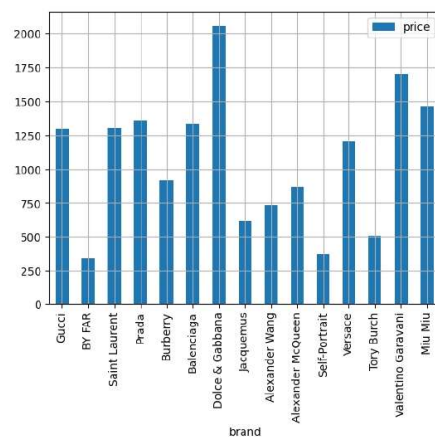
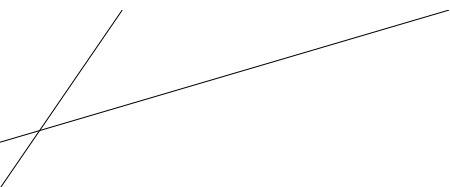
Numero di borse per marchio - sotto gli 800€



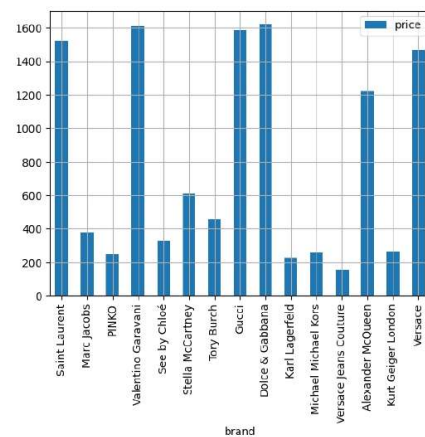
Numero di borse per marchio - sotto i 1500€



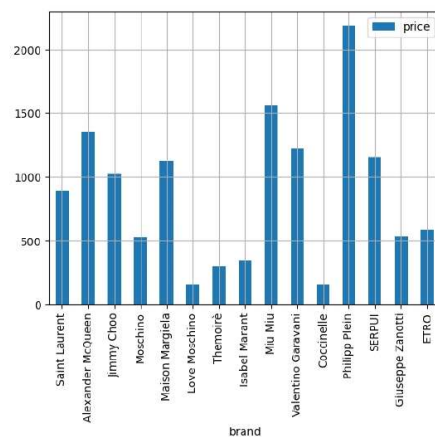
Numero di borse per marchio - sopra i 1500€



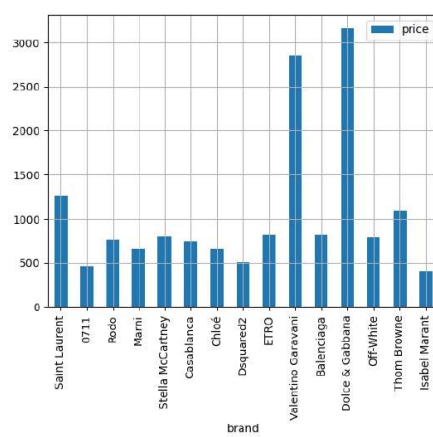
Prezzo medio per brand - Borse mini



Prezzo medio per brand - Borse a tracolla



Prezzo medio per brand - Borse Clutch



Prezzo medio per brand - Borse da mare

## RISULTATI TABELLE

e0

1. Dai primi grafici possiamo dedurre che: le borse a tracolla, le borse clutch e le borse a secchiello hanno un alto numero di borse economiche. Un alto numero di borse mid-range è presente nelle borse da mare e le borse mini. Tra le borse a spalla invece sono relativamente più diffuse le borse sopra i 1500€ indicando come sia il genere di borsa dove i brand di lusso operano maggiormente.
2. I grafici a dispersione offrono una visione più completa del mercato, indicando in quale range di prezzo ogni brand opera.
3. Generando degli istogrammi con il prezzo medio per tipologia di borsa per brand è possibile, non solo confrontare il prezzo medio tra i diversi brand, ma anche il prezzo medio dello stesso brand tra tipologie diverse.
4. Sotto gli 800€ i brand più popolari sono BY FAR, Marc Jacob, Jacquemus e Coccinelle. Questi marchi sono molto popolari in quella fascia di prezzo ma sono quasi non presenti nei range superiori. Al contrario, brand come Saint Laurent, avendo prodotti che si collocano in un range di prezzo molto ampio, è tra il brand più popolari in tutte le fasce di prezzo (11° sotto 800€, 1° sotto i 1500€, 1° sopra i 1500€).
5. Brand come Valentino Garavani, Prada, Gucci sono presenti quasi esclusivamente nella categoria sopra i 1500€, ciò dimostra il diverso target dei vari brand

## Diapositiva 23

---

e0

Questa parte puoi dirla a voce mentre fai vedere i grafici

emanuele.balloni@unimc.it; 2022-09-27T16:00:12.549

## CONCLUSIONI E SVILUPPI FUTURI

- I dati estrapolati sono stati utilizzati per mostrare uno spaccato della distribuzione delle borse sia per tipologia che per prezzo.
- Questi dati possono risultare estremamente utili sia per i gestori del sito Farfetch che per i competitors per elaborare strategie consone con i risultati.
- Le stesse aziende produttrici di borse possono utilizzare questi dati per comprendere meglio il loro posizionamento nel mercato
- Partendo da questo progetto è possibile fare confronti del mercato in due momenti diversi per vederne i trend.
- Applicare la stessa ricerca ad un sito competitor ed analizzarne le differenze



GRAZIE PER L'ATTENZIONE

e0

e1



## Diapositiva 25

---

**e0**

Relatore

emanuele.balloni@unimc.it; 2022-09-27T16:00:59.608

**e1**

Relatore e correlatore vanno nella prima slide, qua scrivi Grazie per l'attenzione o una cosa simile

emanuele.balloni@unimc.it; 2022-09-27T16:06:37.326