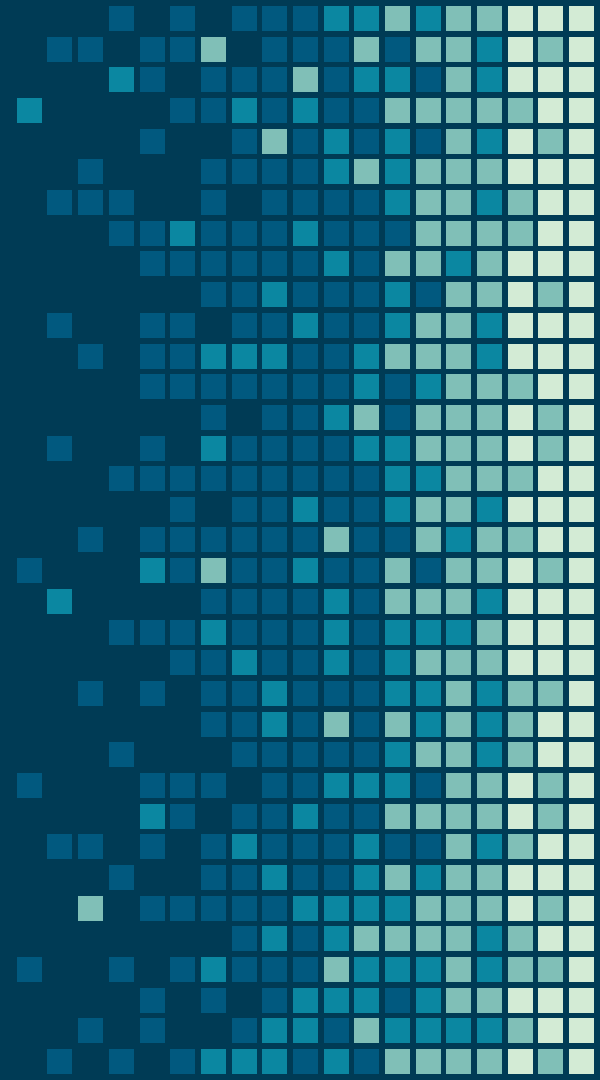


# PROFILAZIONE COMPETITOR ELETTRONICA DI CONSUMO

AIRAGHI DAVIDE

Master Business Intelligence and Big Data Analytics  
Università degli studi Milano Bicocca  
a.a 2020 - 2021



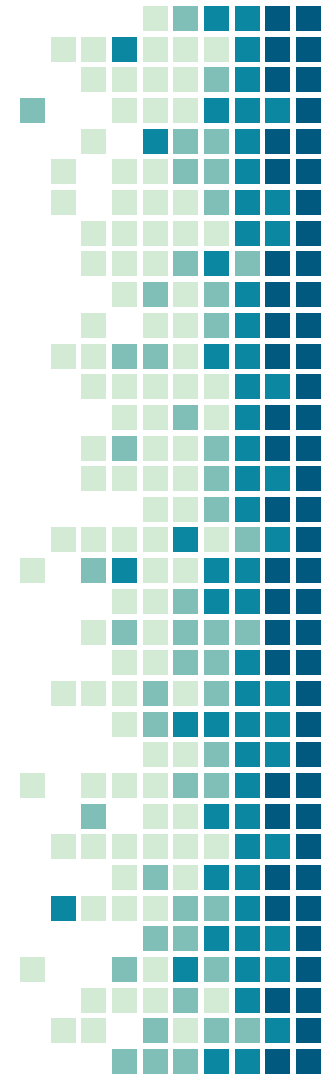
# Agenda

- 1) **CONTESTO DI RIFERIMENTO**
- 2) **OBIETTIVO**
- 3) **ARCHITETTURA**
- 4) **DONE / TO DO**

# Contesto

## **ELETTRONICA DI CONSUMO ON LINE**

Il mercato globale dell' elettronica di consumo é in continua crescita e nel 2020 crescerà del 2.5% rispetto all' anno precedente con un giro di affari che raggiungerà 1.05 trilioni di euro a livello globale con comparti di Piccoli elettrodomestici (+8%), Telecomunicazioni (+3%) e Grandi elettrodomestici (+2%) che faranno da traino.



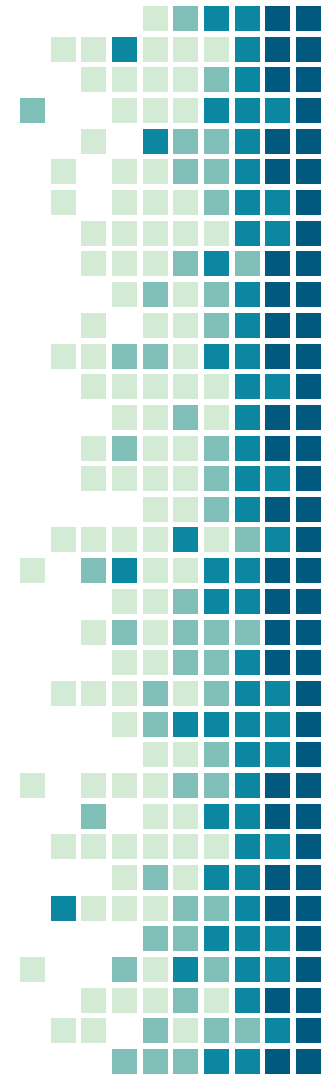
# Players

Il mercato dell' e commerce conta numerosi player che dispongono di negozi fisici oltre che la parte on line.

Per questa prima parte di progetto analizzeremo:

❖ **MEDIAWORLD**

❖ **UNIEURO**



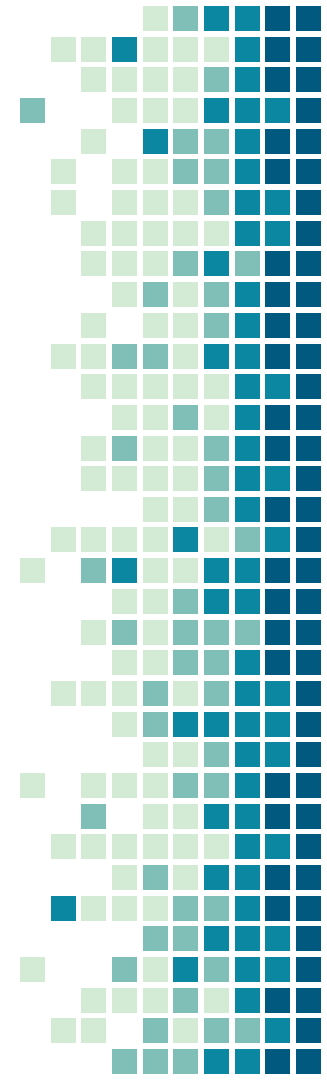
# Obiettivo

L'obiettivo principale del progetto è creare uno strumento che permetta di effettuare una profilazione dei competitor all' interno del perimetro delle vendite on line (e-commerce).

Una volta effettuata la profilazione il manager è in grado di fare una valutazione del sua strategia rispetto agli altri player del settore e implementare azione correttive o migliorative per aumentare l' efficacia della sua proposta.

Le aree tematiche su cui è possibile agire sono:

- conformazione portafoglio prodotti
- scontistica applicata
- target clienti (per fasce di prezzo)
- comportamento durante eventi speciali (Cyber Monday)
- offerta in linea con i trend di mercato

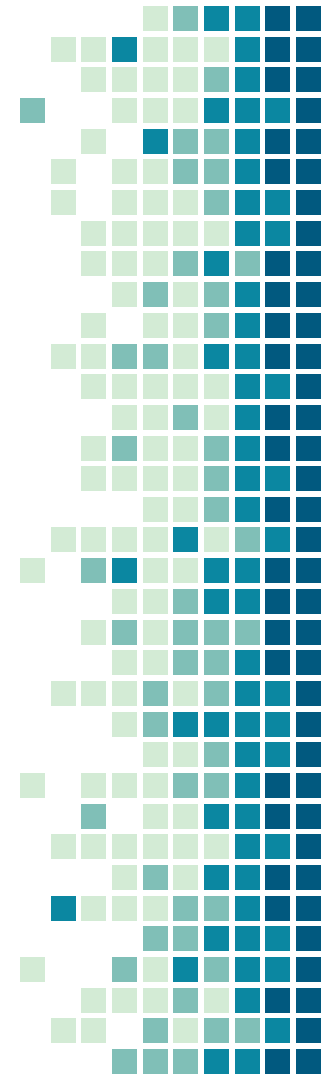


# Tasks

Vista la numerosità dei domini presenti in un sito di e-commerce ci concentriamo per il momento su il dominio degli smartphone.

I task da eseguire sono:

1. Individuazione del perimetro di dati propedeutici per l'analisi
2. Estrazione dei dati dal sito dei competitors
3. Profilazione e pulizia del dataset ottenuto
4. Trasformazione del dataset ottenuto
5. Creazione del cruscotto direzionale per effettuare la comparazione tra competitor



# Architettura



## 1° - Scraping

Utilizzo Selenium per effettuare i task di scraping tramite il browser Google Chrome.

I task sono sviluppati con Python mediante Jupyter Notebook

## 2° - Data profiling

La profilazione del dato ottenuto avviene tramite il tool OpenRefine.

Questa operazione comprende anche l'allineamento tra gli attributi delle diverse sorgenti

## 3° - Data transforming / enrichment

Questo step del processo prevede di aggiungere attributi propedeutici alla profilazione al dataset di partenza

## 4° - Data Viz

L'ultimo step del processo prevede di fruire del dato acquisito tramite dashboard direzionale Power BI

# Scraping

La fase di scraping viene eseguita tramite Selenium e avviene tramite due macro task:

- acquisizione dei link per dei singoli prodotti per ogni pagina presente della categoria (smartphone)
- Utilizzo del link acquisito nello step precedente per recuperare le specifiche di ogni singolo prodotto

Tutti i prodotti disponibili vengono acquisiti

Ho deciso di concentrarmi su due vendor principali Unieuro e MediaWorld:



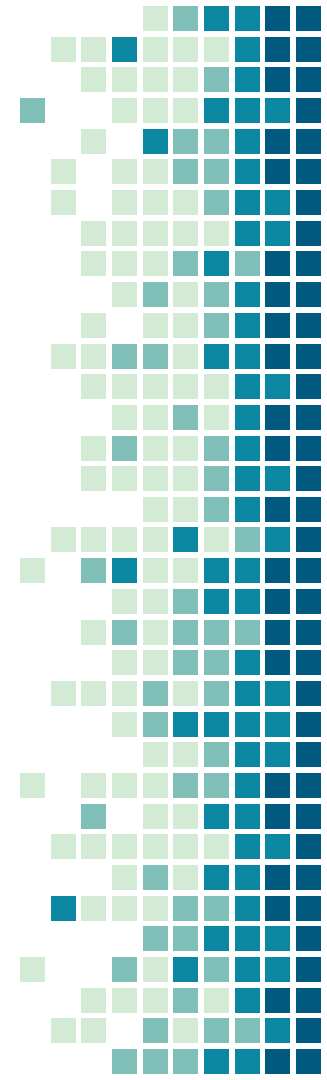
# MediaWorld – caratteristiche e problemi

Il sito di MediaWorld non risulta particolarmente complesso almeno per quanto concerne la sezione smartphone

Gli unici punti di attenzione sono:

- gli attributi delle specifiche tecniche hanno degli spazi → trim a destra e a sinistra della stringa
- Non è presente la marca nella lista delle specifiche ma il pattern della descrizione estesa del prodotto è sempre: «MARCA modello» quindi in questo caso vado a splittare la stringa.
- Sono a volte presente dei duplicati nelle diverse pagine del sito
- le dimensioni del prodotto (altezza, lunghezza, larghezza) sono contenute in un' unica stringa nel formato L x W x H. E' necessario quindi splittare la stringa sfruttando il separatore «x».

**In fase di confronto dei siti si nota che le unità di misura sono differenti rispetto agli altri competitor**



# Unieuro – caratteristiche e problemi

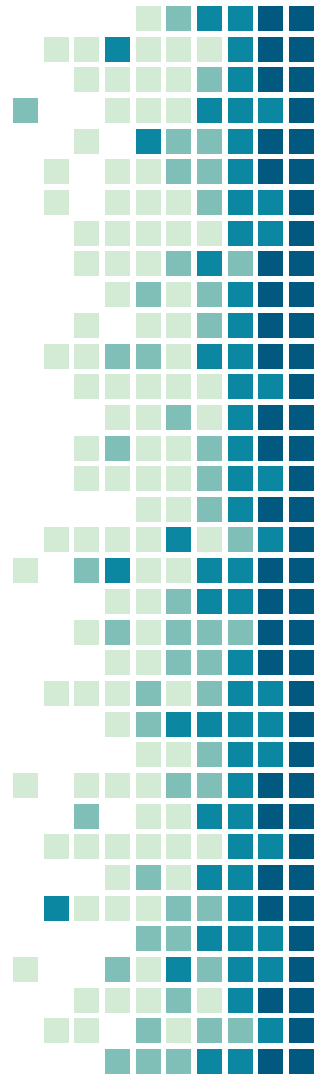
Il sito di Unieuro risulta ben strutturato per quanto riguarda la parte smartphone ma ha una sezione a parte per quanto riguarda gli Iphone pertanto bisogna predisporre due fasi di scraping per questo sito.

Altre problematiche:

Brand, modello e operatore sono da recuperare in modo differente:

Tutte queste informazioni infatti sono contenute nella descrizione estensiva del prodotto (p.e. SAMSUNG Galaxy Note10 Lite Black WindTre) e ho utilizzato delle espressioni regolari in modo da intercettare la maggior parte dei casi

- BRAND: lo troviamo sempre all' inizio della stringa ma a volte viene prima specificato dopo l' operatore → Rimuovo quindi eventuali operatori
- MODELLO: viene specificato dopo il brand e prima di alcuni caratteri/stringhe quali: «,» , «:» , «numero GB» , «numero cm» etc.
- OPERATORE: esiste un' area in cui questo attributo viene specificato ma risulta più comodo recuperarlo dalla descrizione per evitare ulteriori lavorazioni



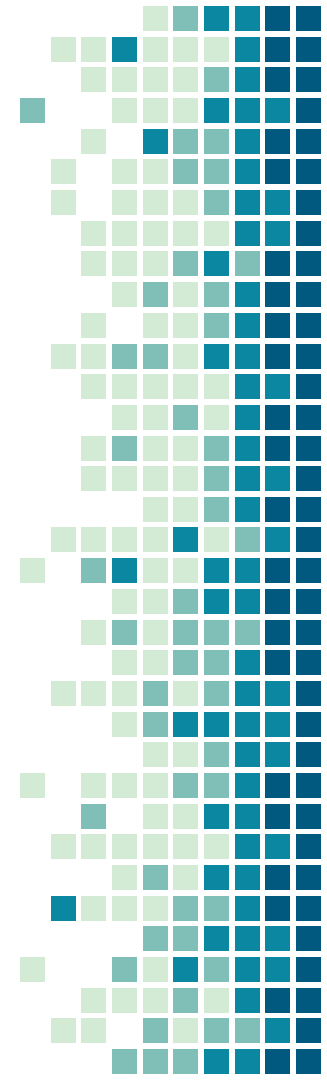
# Data Profiling



La fase di data profiling è stata fatta con OpenRefine e la criticità maggiore è relativa alla riconciliazione degli attributi che in questo caso vengono da fonti eterogenee seppur appartenenti allo stesso dominio.

In particolare per gli attributi:

- **color:** è specificato in inglese sul sito di MediaWorld e in italiano sul sito di Unieuro
- **weight,length,width,height:** hanno unità di misure differenti tra loro
- **display\_type:** ci sono alcuni valori che hanno lo stesso significato ma sono espressi in maniera diversa (SAMOLED = Super Amoled)
- **os:** stessi valori ma scritti in maniera differente
- **brand:** stessi valori ma scritti in maniera differente
- **carrier:** H3G è ora WindTre



# Data Transforming / Data Enrichment

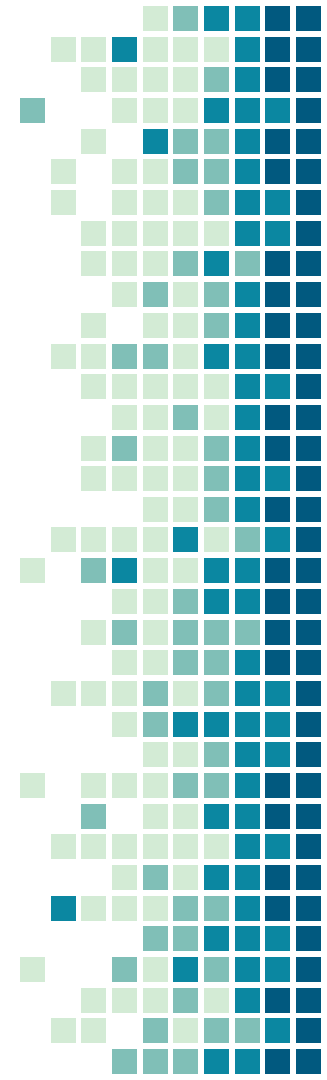
Dopo la fase di data profiling e data cleaning svolta con OpenRefine ho modificato il dataset come segue:

- L'attributo `original_price` assume lo stesso valore di `price` se mancante (significa che non è in sconto)
- Calcolo del risparmio per singolo prodotto ( $\text{price} - \text{original\_price}$ )
- Calcolo del valore percentuale risparmiato per singolo prodotto
- Creazione del flag per identificare i prodotti in sconto
- Assegnazione range di prezzo ai prodotti
- Assegnazione fasce di sconto

# Struttura del dataset ottenuto



campo	datatype	descrizione	unità di misura
tms_loading	timestamp	timestamp del caricamento	--
date_loading	date	data del caricamento	--
os	string	sistema operativo	--
brand	string	marca dello smartphone	--
model	string	modello dello smartphone	--
price	number	prezzo di vendita attuale	€
original_price	number	prezzo originale	€
ram	number	ram	GB
memory	number	memoria interna	GB
sim	string	sim/doppia sim	--
gps	string	gps	--
a-gps	string	assisted gps	--
color	string	colore	--
display_type	string	tipo di display	--
display_size	number	dimensione display	pollici
carrier	string	operatore telefonico	--
battery_capacity	number	capacità della batteria	mAh
height	number	altezza	mm
width	number	larghezza	mm
length	number	lunghezza	mm
wieigth	number	peso	g
price_range	number	fascia di prezzo del prodotto	--
discount	number	valore scontato	€
discount_perc	number	percentuale applicata sconto	--
flag_discount	string	prodotto in sconto / non in sconto	--
discount_range	number	fascia valore scontato	--



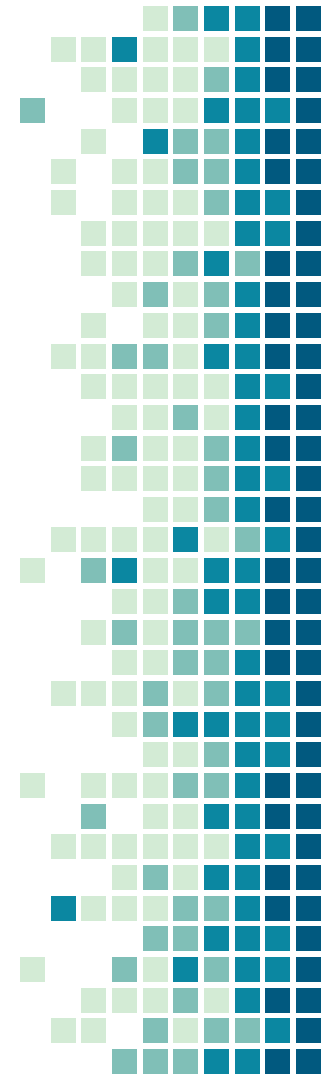
# Data Visualization

Il dataset ottenuto viene importato all'interno di Power BI che permette all'utente di ottenere degli insights sui diversi aspetti riguardanti i propri competitor.

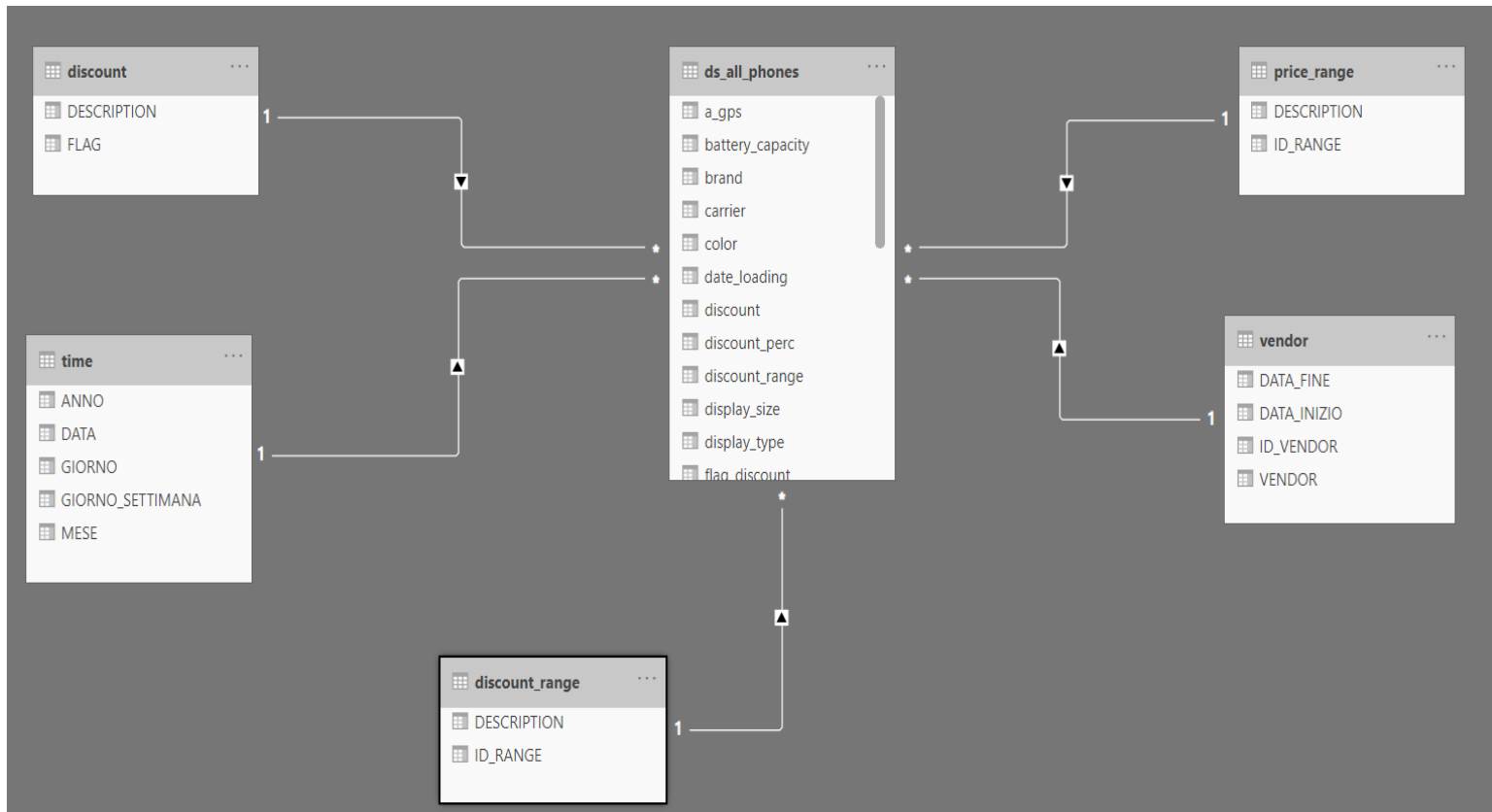
Così facendo l'utente è nella condizione di confrontare la sua strategia e, nel caso, implementare azioni correttive o migliorative per ottenere un vantaggio competitivo.

Aree tematiche:

- Tecnologia: l'utente deve chiedersi se il portafoglio prodotti è in linea con i trend tecnologici del momento p.e: i telefoni dual sim sono in crescita?
- Prezzo: i competitor utilizzano una strategia aggressiva di prezzo rispetto a me? Qual è il target di clienti? Come è distribuita la scontistica dei prodotti? Esiste una tendenza ricorrente nell'applicazione delle promozioni?

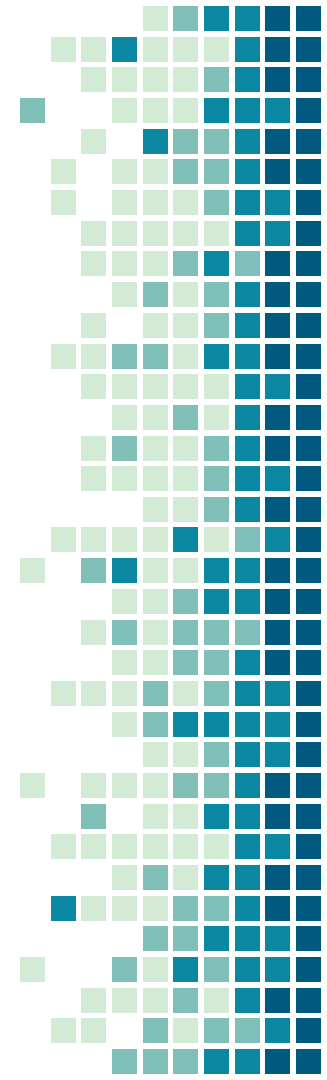


# Data Model



# Insights

- Numerosità portafoglio prodotti differente rispetto i due players
- Unieuro offre più una gamma di brand più ampia rispetto a MediaWorld
- Il dual sim ha ormai la maggioranza del mercato
- Per entrambi i players la fascia di prezzo che interessa più prodotti è quella con prezzo < 200 euro
- Gli smartphone carrier less sono la maggior parte dei prodotti venduti probabilmente per non precludere l'acquisto con il vincolo dell'operatore
- Il portafoglio prodotti di Unieuro è più numeroso (con prodotti di fascia sotto ai 500 euro) e questo permette ad Unieuro di attuare una politica di scontistica aggressiva relativamente al prezzo di questi prodotti.
- Nel periodo di misurazione non si segnalano modifiche al trend della scontistica applicata e il motivo potrebbe essere che non ci sono state campagne di sconto durante quelle giornate





# DONE / TO DO

## DONE

- Raccolta dati dei vendor
- Trasformazione e pulizia del dato
- Riconciliazione del dato
- Creazione dashboard direzionale

## TO DO

- Architettura per la storicizzazione del dato
- Ampliare la profilazione su altre macro categorie di prodotti

GRAZIE!