

**UNIVERSITA' DEGLI STUDI DI TORINO**  
**MASTER DI I LIVELLO IN**  
**ANALISI DATI PER LA BUSINESS INTELLIGENCE E DATA SCIENCE**  
**ANNO ACCADEMICO 2018/2019**

**CANDIDATO: LUIGI CAPPELLETTO**

**DIMOCard loyalty program: processo ETL su database clienti e primi approcci al clustering.**

**ABSTRACT**

Il project work condotto presso l'azienda Dimo SpA ha avuto come scopo la messa in pratica dei principi di base della gestione e analisi dei dati, con particolare attenzione al processo di raccolta, trasformazione ed elaborazione al fine di avere a disposizione un set di dati su cui applicare degli algoritmi di clustering.

I dati utilizzati provengono dal programma fedeltà che l'azienda Dimo SpA ha in essere con i suoi clienti e grazie al quale ha potuto negli anni raccogliere informazioni relative ai titolari della carta fedeltà DIMOCard e soprattutto mantenere uno storico degli acquisti effettuati dai clienti stessi.

Il project work è iniziato con l'analisi dell'architettura del sistema informativo dell'azienda, proseguendo con l'illustrazione delle procedure di estrazione dei dati dai sistemi ERP AS400 e di Marketing Automation Acoustic Campaign, sino ad arrivare alla predisposizione del file contenente tutti i dati necessari agli algoritmi di clustering.

Utilizzando il linguaggio Python, sono state applicate tecniche di data cleaning, data filtering, data transformation, e data normalization. Al fine di individuare il numero di cluster più adatto è stato utilizzato lo studio dei dendogrammi.

Si è proceduto infine ad analizzare il risultato dell'applicazione dell'algoritmo di clustering k-means evidenziando le caratteristiche salienti dei raggruppamenti ottenuti al fine di poterli utilizzare successivamente nella campagne di marketing.

L'ultima analisi riguarda invece i possibili sviluppi futuri dell'analisi dati all'interno dell'azienda Dimo SpA con particolare attenzione all'implementazione di un data warehouse presso il quale raccogliere in modo organizzato e storicizzato tutti i dati che, previa l'utilizzo di sistemi di analisi e di data visualization, diventeranno il principale supporto alle attività decisionali del management.