



## Data Science Challenge 2022

*Deloitte for UNIPD*

11 marzo 2022

# Piacere di conoscervi!



**Sergio Pontello**  
Senior Manager

Sergio ha più di 10 anni di esperienza su progetti di Analytics, implementando e gestendo soluzioni di Advanced Analytics, AI e Business Intelligence. Ha esperienza nel guidare gli utenti di business in progetti di trasformazione strategica e nel supportare i processi decisionali con un approccio data-driven. Si è laureato in matematica a Milano con una specializzazione in geometria algebrica. Sergio ha diverse esperienze in progetti nazionali e internazionali gestendo team grandi e complessi. I principali settori in cui Sergio è stato coinvolto sono: FSI, Energia e Utilities, Automotive, Retail.



**Luca Grandinetti**  
Data Scientist

Luca è un data scientist che si è specializzato in ambito customer analytics collaborando a stretto contatto con le divisioni marketing dei clienti per i quali ha lavorato progettando e sviluppando una molteplicità di modelli di machine learning quali: CB Clustering by Price Sensitivity, Propensione al fotovoltaico, Lead engine per stazioni di ricarica per B2B, Churn Propensity Model, SMS classifier etc... Il suo background nella fisica dei sistemi complessi gli fornisce un'ampia varietà di competenze quantitative.



**Sara Ceschin**  
Data Scientist

Sara lavora in Deloitte come data scientist e data analyst. Ha conseguito la laurea magistrale in Scienze Statistiche all'Università di Padova nel 2019. Nel 2020 è entrata in Deloitte e si occupa di data modeling e machine learning. Lavora principalmente sul mondo automotive e del credito

## DELOITTE AI & DATA OFFERING

### Big Data & Insights driven Organization

Big data to answer critical and complex business questions



An insight-driven organization embeds analysis, data, and reasoning into the decision making process, every day.

### AI, Cognitive & Advanced Analytics

Analytics in the IOT improve processes efficiency and optimize assets usage



Cognitive technologies automatically learn autonomously and extend both human and machine actions

Advanced analytical methods provide insurers opportunities to obtain client-specific insights,

### A&C GLOBAL

29

GLOBAL DELIVERY CENTERS

28

DIGITAL STUDIOS

28

GREENHOUSES

15.000+

GLOBAL HEADCOUNT

### A&C ITALY

250+

RESOURCES IN ITALY

30+

BROAD TECH SKILLS

50+

CLIENTS

# Il contesto

L'aumento della diffusione degli **open data** negli ultimi anni li sta rendendo sempre più uno strumento utile per le aziende perché permette di integrare la propria visione e prendere decisioni supportate dai dati.

Dal 2015 il *Ministero dello Sviluppo Economico* mette a disposizione i **prezzi del carburante** nelle varie stazioni di rifornimento sul territorio italiano.

# Il contesto

L'aumento della diffusione degli **open data** negli ultimi anni li sta rendendo sempre più uno strumento utile per le aziende perché permette di integrare la propria visione e prendere decisioni supportate dai dati.

Dal 2015 il *Ministero dello Sviluppo Economico* mette a disposizione i **prezzi del carburante** nelle varie stazioni di rifornimento sul territorio italiano.



## **Un nuovo player nella vendita del carburante si sta avvicinando al mercato italiano**

Il cliente ha chiesto a Deloitte di studiare le strategie di prezzo dei competitor, individuare i fattori che contribuiscono maggiormente alla definizione del prezzo (a meno di fattori globali) dei singoli impianti e suggerire delle possibili price strategy per impianti di nuova realizzazione

# Il contesto

L'aumento della diffusione degli **open data** negli ultimi anni li sta rendendo sempre più uno strumento utile per le aziende perché permette di integrare la propria visione e prendere decisioni supportate dai dati.

Dal 2015 il *Ministero dello Sviluppo Economico* mette a disposizione i **prezzi del carburante** nelle varie stazioni di rifornimento sul territorio italiano.



## **Un nuovo player nella vendita del carburante si sta avvicinando al mercato italiano**

Il cliente ha chiesto a Deloitte di studiare le strategie di prezzo dei competitor, individuare i fattori che contribuiscono maggiormente alla definizione del prezzo (a meno di fattori globali) dei singoli impianti e suggerire delle possibili price strategy per impianti di nuova realizzazione



## **Obiettivo**

Analizzare i prezzi delle stazioni di riferimento per comprendere se ci sono dei fattori che influenzano i prezzi e quali sono (al netto di fattori globali e macroeconomici)

# Il contesto

L'aumento della diffusione degli **open data** negli ultimi anni li sta rendendo sempre più uno strumento utile per le aziende perché permette di integrare la propria visione e prendere decisioni supportate dai dati.

Dal 2015 il *Ministero dello Sviluppo Economico* mette a disposizione i **prezzi del carburante** nelle varie stazioni di rifornimento sul territorio italiano.



## Un nuovo player nella vendita del carburante si sta avvicinando al mercato italiano

Il cliente ha chiesto a Deloitte di studiare le strategie di prezzo dei competitor, individuare i fattori che contribuiscono maggiormente alla definizione del prezzo (a meno di fattori globali) dei singoli impianti e suggerire delle possibili price strategy per impianti di nuova realizzazione



## Obiettivo

Analizzare i prezzi delle stazioni di riferimento per comprendere se ci sono dei fattori che influenzano i prezzi e quali sono (al netto di fattori globali e macroeconomici)



## Cosa valutiamo?

- Originalità ed efficacia della soluzione proposta
- Chiarezza e creatività della presentazione
- Approccio al problema
- Bonus per ulteriori analisi



## Se avete ancora dubbi

- Sessione di Q&A
- Docenti Tutor Università
- Sergio Pontello, Luca Grandinetti, Sara Ceschin (Deloitte)

# La challenge



## ANALISI ESPLORATIVA:

- **Preprocessing & data cleaning:** gli open data devono essere puliti, individuare eventuali duplicati, inconsistenze e valori nulli e gestire tali problemi al fine di consolidare il dato per le analisi successive
- **Analizzare i prodotti venduti,** individuarne uno per le analisi successive che si ritiene di maggior interesse
- **Osservare il prezzo medio nazionale** del prodotto scelto nel tempo, fornire una **descrizione qualitativa dell'andamento**. Quali dati potrebbero aiutare a predirne il trend? (Non è richiesta la realizzazione di un modello di forecast.)

# La challenge



## ANALISI ESPLORATIVA:

- **Preprocessing & data cleaning:** gli open data devono essere puliti, individuare eventuali duplicati, inconsistenze e valori nulli e gestire tali problemi al fine di consolidare il dato per le analisi successive
- **Analizzare i prodotti venduti,** individuarne uno per le analisi successive che si ritiene di maggior interesse
- **Osservare il prezzo medio nazionale** del prodotto scelto nel tempo, fornire una **descrizione qualitativa dell'andamento**. Quali dati potrebbero aiutare a predirne il trend? (Non è richiesta la realizzazione di un modello di forecast.)



## MODELLAZIONE:

- Il prezzo di ogni distributore è influenzato da fattori macroeconomici esterni e dalle caratteristiche statiche e intrinseche del distributore. **Elaborare una metrica basata sul prezzo** del distributore che non sia correlata ai trend macroeconomici ma solamente **sulle caratteristiche intrinseche del distributore**.
- Le caratteristiche intrinseche e statiche del distributore, **come influenzano il prezzo?**



# La challenge



## ANALISI ESPLORATIVA:

- **Preprocessing & data cleaning:** gli open data devono essere puliti, individuare eventuali duplicati, inconsistenze e valori nulli e gestire tali problemi al fine di consolidare il dato per le analisi successive
- **Analizzare i prodotti venduti,** individuarne uno per le analisi successive che si ritiene di maggior interesse
- **Osservare il prezzo medio nazionale** del prodotto scelto nel tempo, fornire una **descrizione qualitativa dell'andamento**. Quali dati potrebbero aiutare a predirne il trend? (Non è richiesta la realizzazione di un modello di forecast.)



## MODELLAZIONE:

- Il prezzo di ogni distributore è influenzato da fattori macroeconomici esterni e dalle caratteristiche statiche e intrinseche del distributore. **Elaborare una metrica basata sul prezzo** del distributore che non sia correlata ai trend macroeconomici ma solamente **sulle caratteristiche intrinseche del distributore**.
- Le caratteristiche intrinseche e statiche del distributore, **come influenzano il prezzo?**

## BONUS

- Integrare l'analisi con dati esterni (e.g. Istat, prezzo del petrolio...)
- Analizzare gli effetti globali e macroeconomici che influenzano il prezzo del carburante
- Suggerire possibili aree di interesse per la costruzione di nuove stazioni di rifornimento

# I dati

## ANAGRAFICA IMPIANTI

ID IMPIANTO	Codice di identificazione
GESTORE	Ragione sociale dell'impresa
BANDIERA	Insegna del distributore
TIPO IMPIANTO	Tipologia di strada su cui si trova l'impianto (autostrada, strada statale, altro)
NOME IMPIANTO	Nome
INDIRIZZO	Indirizzo
COMUNE	Comune
PROVINCIA	Provincia
LATITUDINE	Coordinata espressa in gradi decimali
LONGITUDINE	Coordinata espressa in gradi decimali

## PREZZO ALLE 8

ID IMPIANTO	Codice di identificazione
DESCCARBURANTE	Tipo di carburante (benzina, gasolio, GPL, ...)
PREZZO	Prezzo €/litro
ISSELF	Modalità di servizio (1=self service, 0=servito)
DTCOMU	Data e ora



Link del dataset per tutti :

[dlt4unipdchallenge2022.ml](https://www.mise.gov.it/index.php/it/open-data/elenco-dataset/2032336-carburanti-prezzi-praticati-e-anagrafica-degli-impianti)



Maggiori informazioni sui dati sono presenti sul sito del *Ministero dello Sviluppo Economico*

<https://www.mise.gov.it/index.php/it/open-data/elenco-dataset/2032336-carburanti-prezzi-praticati-e-anagrafica-degli-impianti>

# L'output atteso



## RELAZIONE FINALE

Una relazione finale che descriva

- Le fasi di sviluppo della challenge
- Le analisi effettuate
- I risultati ottenuti



## MATERIALE DI SUPPORTO

- Eventuale materiale integrativo utilizzato per le analisi, ad esempio open data



## CODICE E OUTPUT

- I codici utilizzati (funzionanti!)
- Eventuali grafici e file di output non riportati nella relazione



È possibile usare qualsiasi strumento software e linguaggio

Il materiale finale dovrà essere consegnato via mail a:

- [guidolin@stat.unipd.it](mailto:guidolin@stat.unipd.it)
- [sceschin@deloitte.it](mailto:sceschin@deloitte.it)
- [lgrandinetti@deloitte.it](mailto:lgrandinetti@deloitte.it)
- [spontello@deloitte.it](mailto:spontello@deloitte.it)

Presentazione challenge  
(oggi)

Termine Iscrizioni  
(22-03-2022)

Check Point  
(Fine Aprile)

Consegna  
04-06-2022

# Q&A



# Thank you

This publication contains general information only, and none of the member firms of Deloitte Touche Tohmatsu Limited, its member firms, or their related entities (collective, the "Deloitte Network") is, by means of this publication, rendering professional advice or services. Before making any decision or taking any action that may affect your business, you should consult a qualified professional adviser. No entity in the Deloitte Network shall be responsible for any loss whatsoever sustained by any person who relies on this publication.

As used in this document, "Deloitte" means Deloitte Consulting LLP, a subsidiary of Deloitte LLP. Please see [www.deloitte.com/us/about](http://www.deloitte.com/us/about) for a detailed description of the legal structure of Deloitte USA LLP, Deloitte LLP and their respective subsidiaries. Certain services may not be available to attest clients under the rules and regulations of public accounting.

**Copyright** © 2022 Deloitte Consulting.  
**All rights reserved. Member of Deloitte Touche Tohmatsu Limited**



## **Sergio Pontello**

Senior Manager, Analytics & Cognitive

**Contact:** [spontello@deloitte.it](mailto:spontello@deloitte.it)



## **Luca Grandinetti**

Data Scientist, Analytics & Cognitive

**Contact:** [lgrandinetti@deloitte.it](mailto:lgrandinetti@deloitte.it)



## **Sara Ceschin**

Data Scientist, Analytics & Cognitive

**Contact:** [sceschin@deloitte.it](mailto:sceschin@deloitte.it)

# Google Colab Notebooks



Colab è un ambiente interattivo che permette di scrivere ed eseguire codice (sia Python che R) nel browser come jupyter. Quello che ti serve è un account Google (se non ce l'hai è molto semplice da creare).

- Nessuna configurazione da effettuare
- Facilità di condivisione
- Accesso alla GPUs (non necessaria per questa challenge)



**Puoi usare qualsiasi strumento software. Colab è un semplice consiglio!**