

# DATA SCIENCE 4 CITIZENS

LABORATORIO DI DATA SCIENCE



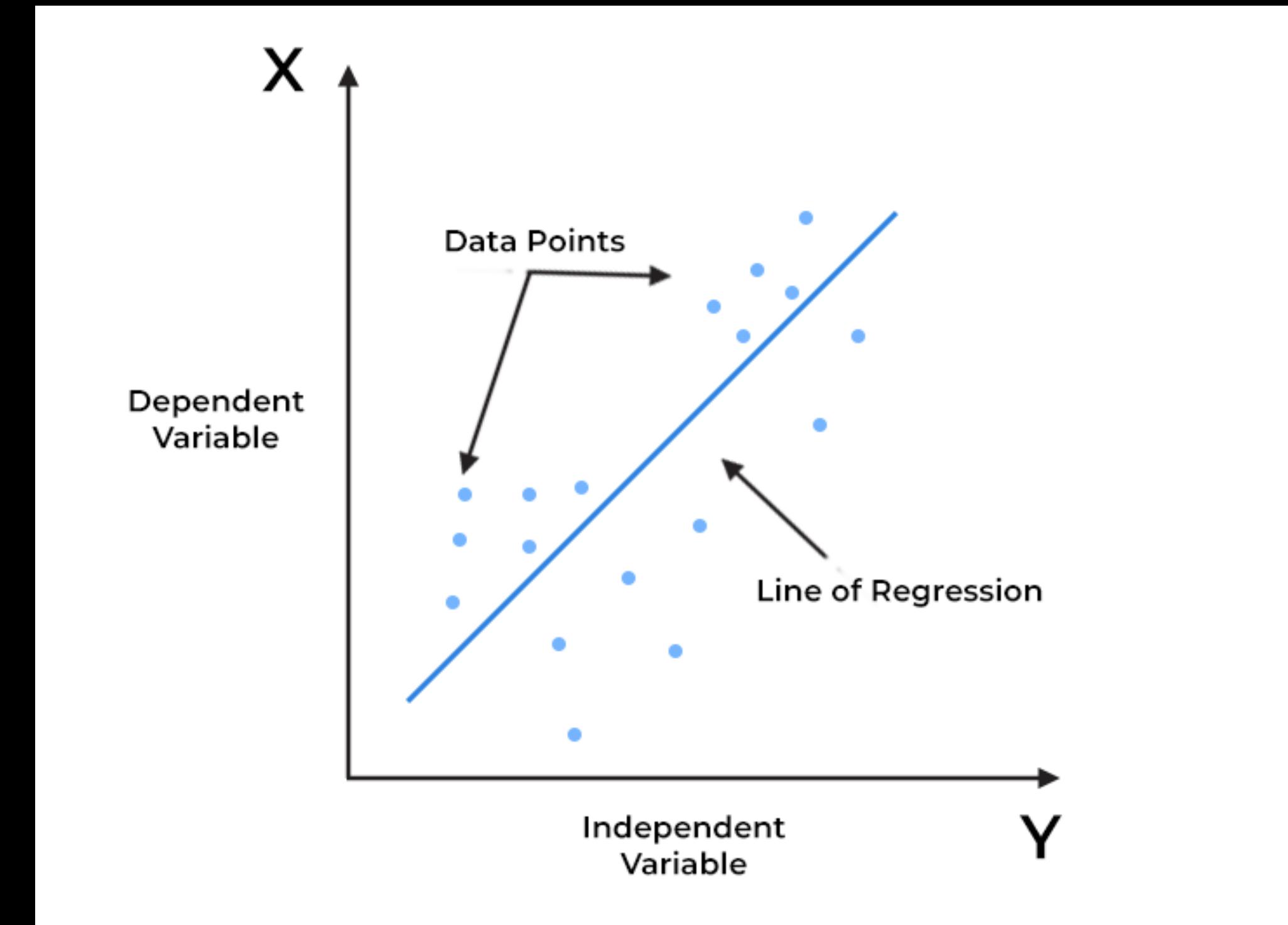
LICEO SCIENTIFICO  
LORENZO RESPIGHI

# REGRESSIONE LINEARE

## COS'E'?

Un metodo statistico che definisce una relazione lineare tra una variabile indipendente ( $x$ ) e la variabile dipendente ( $y$ ).

E' l'algoritmo che usiamo per prevedere la **popolarità delle canzoni** ( $y$ ), date le variabili che abbiamo osservato ( $x$ ).



Tra i principali assunti del modello lineare abbiamo l'assenza di **multicollinearità**.

Si ha *multicollinearità* quando le variabili  $X$  sono tra di loro correlate (ovvero sussiste una relazione di tipo lineare anche tra di loro).

Quando un modello presenta multicollinearità i suoi risultati possono essere "biased", ovvero sistematicamente diversi dai valori reali.



# C'E' MULTICOLLINEARITA'?

## STEP 1

Guardiamo la matrice di correlazione: *possiamo vedere se ci sono variabili correlate tra loro?*

## STEP 2

Utilizzare il **VIF test**: con questo test possiamo verificare se la varianza associata ad una variabile indipendente viene "gonfiata" dalla presenza di multicollinearità.

Se il VIF associato alle variabili è maggiore 5, possiamo considerare il nostro modello "*biased*".



UNIVERSITA' DEGLI STUDI DI MILANO & LICEO SCIENTIFICO L. RESPIGHI



LICEO SCIENTIFICO  
LORENZO RESPIGHI

# CHI E' IL DATA SCIENTIST?

## Professioni più richieste nei prossimi 10 anni

Di seguito una prima tabella con quelle che saranno i lavori più richiesti nel prossimo decennio.

Tipologia di lavoro	Settore
1. Analisti e specialisti in IA e Machine Learning	Tecnologia
2. Specialisti dei Big Data and Cloud	Informatica-tecnologia
3. Consulenti di digital marketing	Marketing-pubblicità
4. Specialisti nella trasformazione digitale	Informatica
5. Project Manager	Business
6. Ingegneri Fintech e legati alla robotica	Finanza e sistemi robotici

# Data Scientist, il professionista più sexy del XXI secolo

*Un po' matematico, un po' informatico. Deve capire di statistica, saper usare un foglio excel ed aiutare l'azienda per la quale lavora a prevedere il futuro. Non è uno sciamano ma un nuovo professionista, lo 'scienziato dei dati'. Che negli Usa è tra i più contesi*

## 10 Lavori Più Richiesti Per I

- Data Analyst / Data Scientist / Cyber Security Analyst
- Java / C++ Developer
- Sviluppatore Python
- Machine Learning / AI Expert
- Digital Project Manager
- Project Engineer Energetico
- Architetto / Ingegnere Edile
- Legal Consultant / Esperti Contabili / Commercialisti
- Social Media Manager & Content Creator



# CHI E' IL DATA SCIENTIST?

Il Data Scientist è un professionista dalle molteplici competenze tecniche:

- Linguaggi di programmazione (R, Python, SQL, Stata, MATLAB, ... e chi più ne ha più ne metta)
- Estrazione dei dati, pulizia dei dati.
- Architettura dei database.
- Analisi dati.
- Inferenza e previsione con modelli di Machine Learning.
- Data Visualization!
- ...
- Competenze specifiche nell'ambito di riferimento.

Ma deve avere anche diverse competenze relazionali e comunicative! (Altrimenti nessuno capirebbe nulla del suo lavoro)



LICEO SCIENTIFICO  
LORENZO RESPIGHI

# IO SONO:

Io sono Doina, ho 24 anni e sono una studentessa dell'ultimo anno del corso in Data Science and Economics presso l'Università degli Studi di Milano.

Il mio percorso :

- Ho fatto il Liceo Linguistico, dove ho studiato Inglese, Tedesco, Russo.
- Mi sono laureata in Economia all'Università di Padova.
- Sto completando la Laurea Magistrale in Data Science and Economics.

# VOI?



LICEO SCIENTIFICO  
LORENZO RESPIGHI

