

Modulo IA

Analisi di Big Data basata su Intelligenza Artificiale

Docente

- **Attualmente**

- Libero professionista e consulente in ambito ICT e Data Science/Analytics
- Docente, formatore, educatore
- Ricercatore indipendente, divulgatore ed attivista

- **25+ anni di carriera in ICT**

- Accademia, R&D, industria, agenzie multilaterali, non-profit
- PM, coordinatore di progetti e team, ricercatore, docente
- Italia ed estero (EU, UK, US, UN)

- **Laurea specialistica (v.o.) e Master di II livello**

- Ingegneria Informatica
- Ingegneria e progettazione per la cooperazione e lo sviluppo

- **Specializzazione e certificazioni**

- Data Science, Humane Technology
- Piattaforme di analytics (KNIME)

Qualche minuto “leggero”...

... sul “peso” odierno di Analisi di Big Data ed Intelligenza Artificiale



Geoffrey Hinton

Facts



Ill. Niklas Elmehed © Nobel
Prize Outreach

Geoffrey Hinton

The Nobel Prize in Physics 2024

Born: 6 December 1947, London, United Kingdom

Affiliation at the time of the award: University of Toronto,
Toronto, Canada

Prize motivation: “for foundational discoveries and
inventions that enable machine learning with artificial
neural networks”

Prize share: 1/2

Work

When we talk about artificial intelligence, we often mean machine learning using artificial neural networks. This technology was originally inspired by the structure of the brain. In an artificial neural network, the brain's neurons are represented by nodes that have different values. In 1983–1985, Geoffrey Hinton used tools from statistical physics to create the Boltzmann machine, which can learn to recognise characteristic elements in a set of data. The invention became significant, for example, for classifying and creating images.

Geoffrey Hinton

Facts



Ill. Niklas Elmehed © Nobel
Prize Outreach

Geoffrey Hinton

The Nobel Prize in Physics 2024

Born: 6 December 1947, London, United Kingdom

Affiliation at the time of the award: University of Toronto,
Toronto, Canada

Prize motivation: “for foundational discoveries and
inventions that enable machine learning with artificial
neural networks”

Prize share: 1/2

Work

When we talk about artificial intelligence, we often mean machine learning using artificial neural networks. This technology was originally inspired by the structure of the brain. In an artificial neural network, the brain's neurons are represented by nodes that have different values. In 1983–1985, Geoffrey Hinton used tools from statistical physics to create the Boltzmann machine, which can learn to recognise characteristic elements in a set of data. The invention became significant, for example, for classifying and creating images.

Demis Hassabis

Facts



Ill. Niklas Elmehed © Nobel
Prize Outreach

Demis Hassabis

The Nobel Prize in Chemistry 2024

Born: 27 July 1976, London, United Kingdom

Affiliation at the time of the award: Google DeepMind,
London, United Kingdom

Prize motivation: “for protein structure prediction”

Prize share: 1/4

Work

Proteins control and drive all the chemical reactions that together are the basis of life. Proteins generally consist of 20 different amino acids. These are linked together in long strings that fold up to make a three-dimensional structure. In 2020, Demis Hassabis and John Jumper presented an AI model called AlphaFold2. With its help, they have been able to predict the structure of virtually all known proteins. AlphaFold2 has been widely used in many areas, including research into pharmaceuticals and environmental technology.

Demis Hassabis

Facts



Ill. Niklas Elmehed © Nobel
Prize Outreach

Demis Hassabis

The Nobel Prize in Chemistry 2024

Born: 27 July 1976, London, United Kingdom

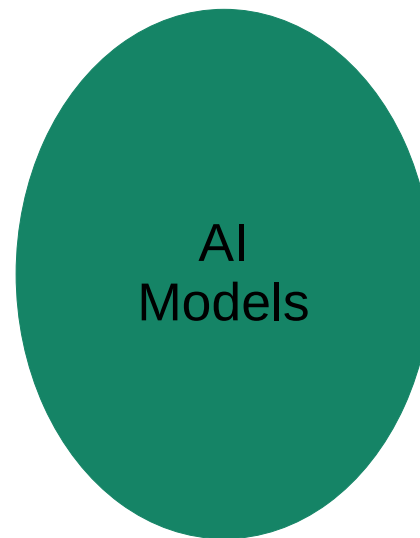
Affiliation at the time of the award: Google DeepMind,
London, United Kingdom

Prize motivation: “for protein structure prediction”

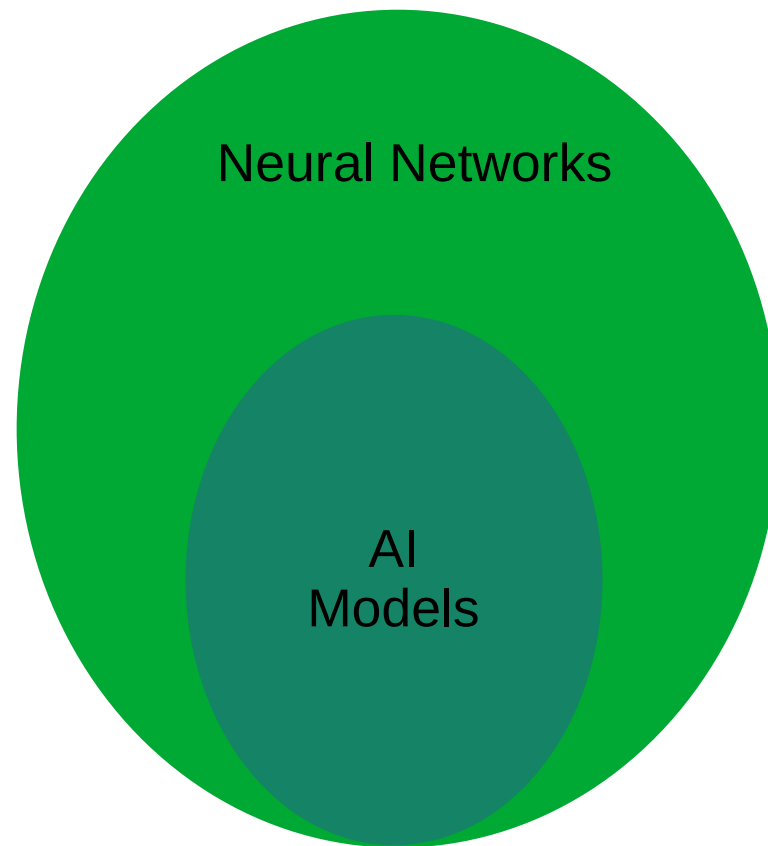
Prize share: 1/4

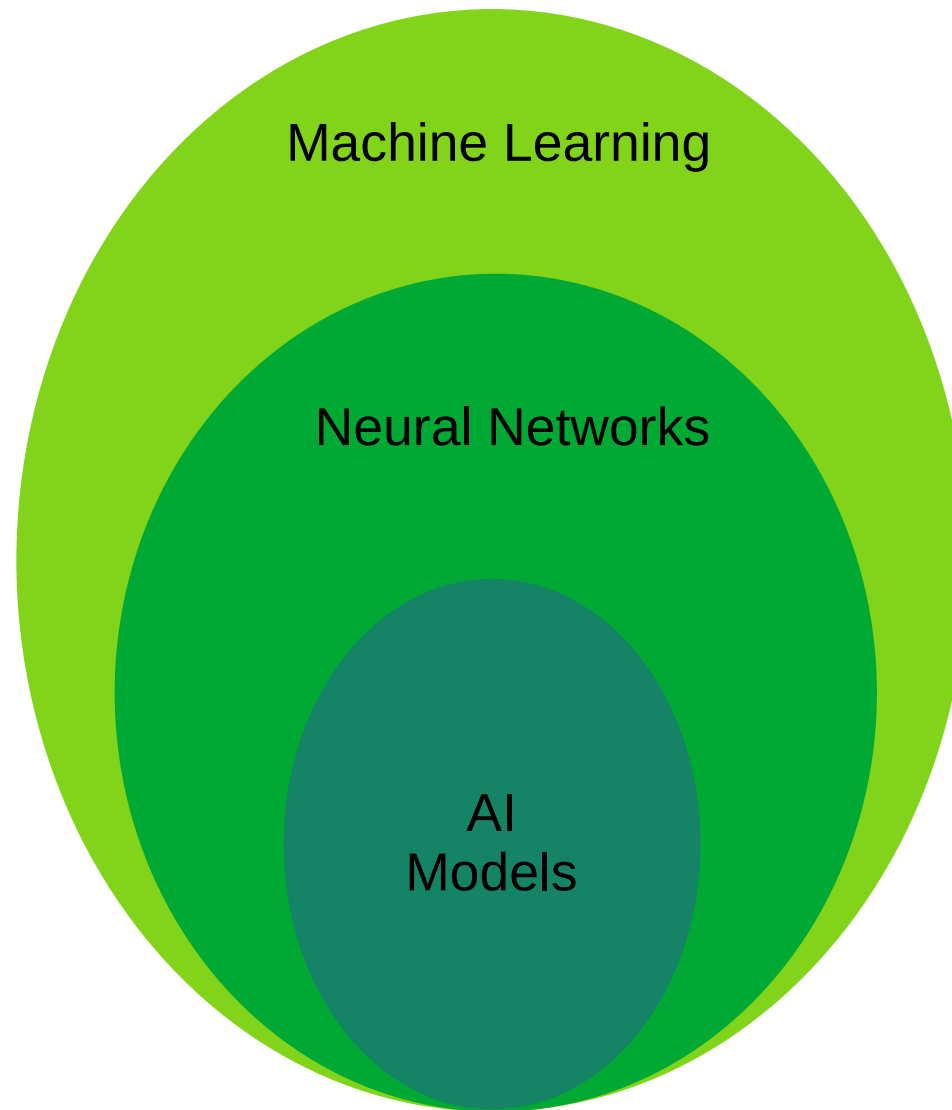
Work

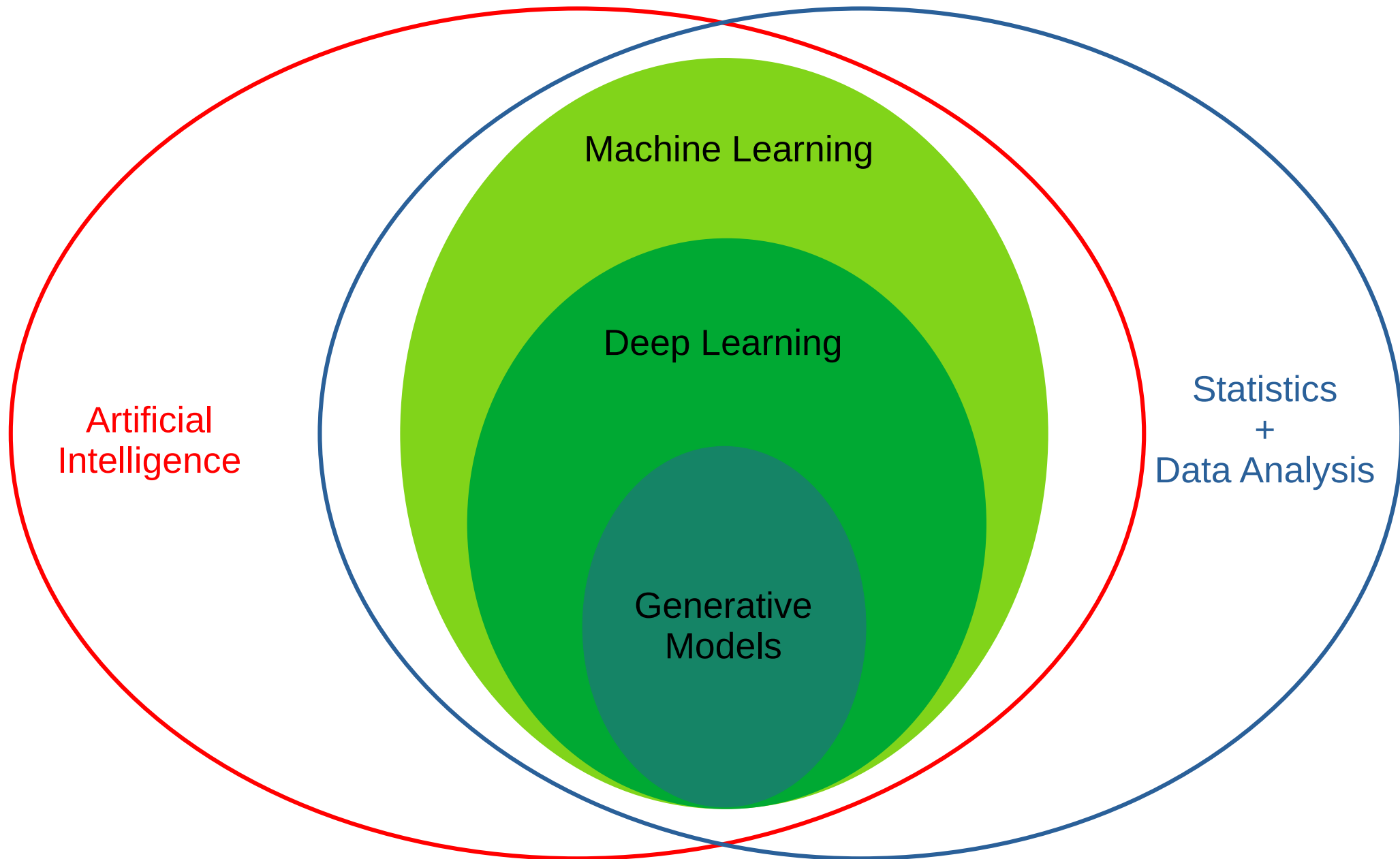
Proteins control and drive all the chemical reactions that together are the basis of life. Proteins generally consist of 20 different amino acids. These are linked together in long strings that fold up to make a three-dimensional structure. In 2020, Demis Hassabis and John Jumper presented an AI model called AlphaFold2. With its help, they have been able to predict the structure of virtually all known proteins. AlphaFold2 has been widely used in many areas, including research into pharmaceuticals and environmental technology.

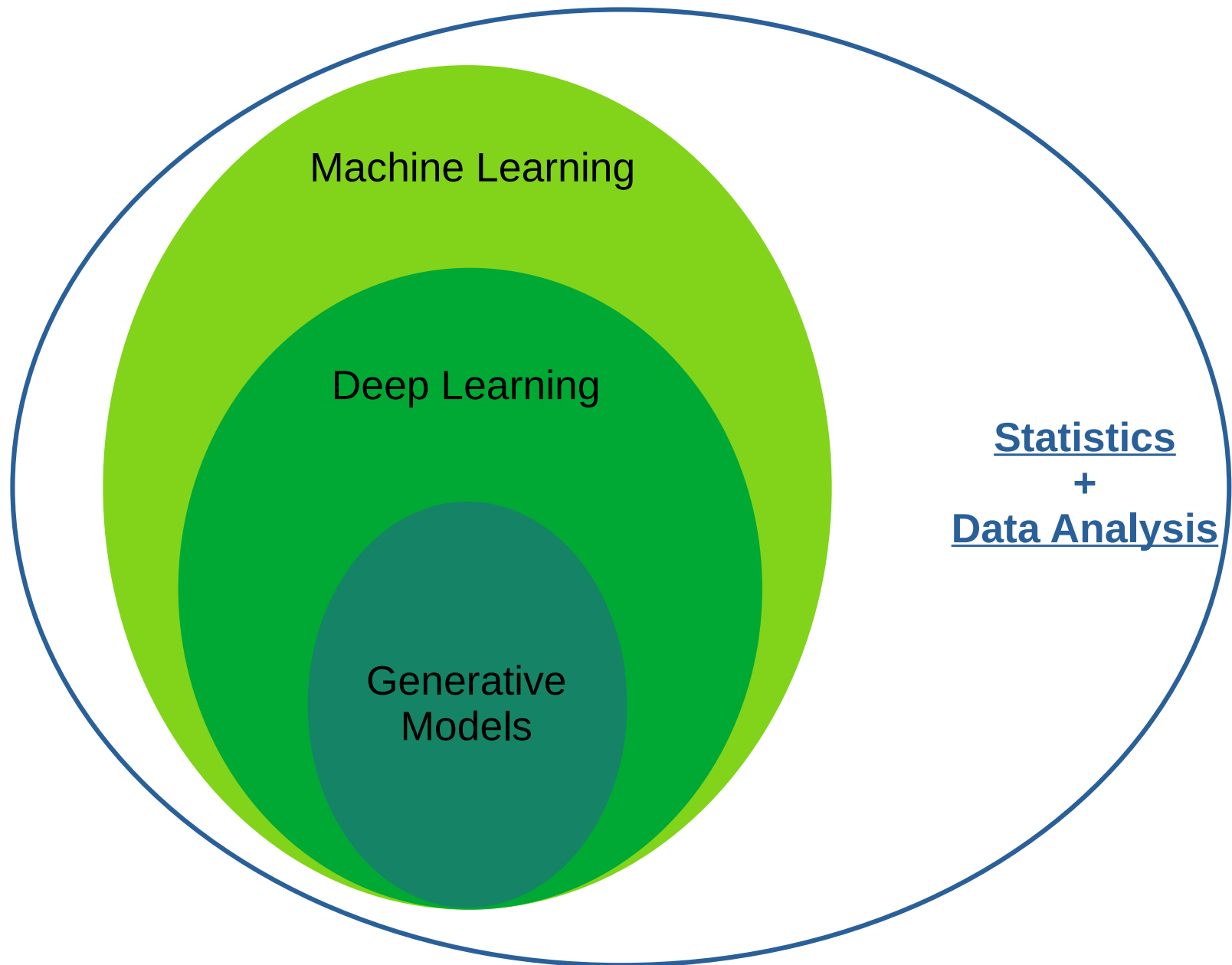


ChatGPT, GPT4, Gemini, Claude, Llama, Grok, DeepSeek, AlphaGo, AlphaZero, AlphaFold(2), DALL-E 3, MidJourney, Stable Diffusion, SORA...













Social Media



Social Media ↔ Economia Digitale



Social Media ↔ Economia Digitale ↔ Big Tech

Machine Learning

Machine Learning ↔ Statistics + Data Analysis

Google

 Apple

 Meta

amazon


 Microsoft

 **nVIDIA®**

The Google logo, consisting of the word "Google" in its signature multi-colored font.The Apple logo, featuring a black silhouette of an apple with a bite taken out of it, followed by the word "Apple" in a black sans-serif font.The Meta logo, featuring a blue infinity symbol followed by the word "Meta" in a bold, black sans-serif font.The Amazon logo, featuring the word "amazon" in a bold, black sans-serif font with a curved orange arrow underneath it.The Microsoft logo, featuring the four-pane Windows logo in red, green, blue, and yellow, followed by the word "Microsoft" in a grey sans-serif font.The NVIDIA logo, featuring a green stylized eye icon followed by the word "NVIDIA" in a bold, black sans-serif font.

Capitalizzazione (totale):
ca. **\$15000B**, 3rd trimestre 2024

Fatturato annuale (individuale):
centinaia di miliardi di dollari

Profitto annuale (individuale):
decine di miliardi di dollari



Pioniere



Indovinate un po' chi ci (ha) lavora(to)?

Google



Google



Machine Learning ↔ **Statistics + Data Analysis**

Cosa imparo nell'unità IA.1?

- **Primi strumenti moderna Data Analysis/Analytics**
 - Tipologie di analisi
 - Descrittiva (→ Analysis)
 - Predittiva/Prescrittiva (→ Analytics)
 - Preparazione dei dati (“Data Wrangling”)
 - Visualizzazione dei dati per primi rapporti statistici
 - “Affaccio” sul Machine Learning
- **L'uso a livello iniziale di alcuni strumenti di lavoro**
 - Python-based (pandas, Matplotlib, seaborn, Plotly)

Cosa imparo nell'unità IA.1?

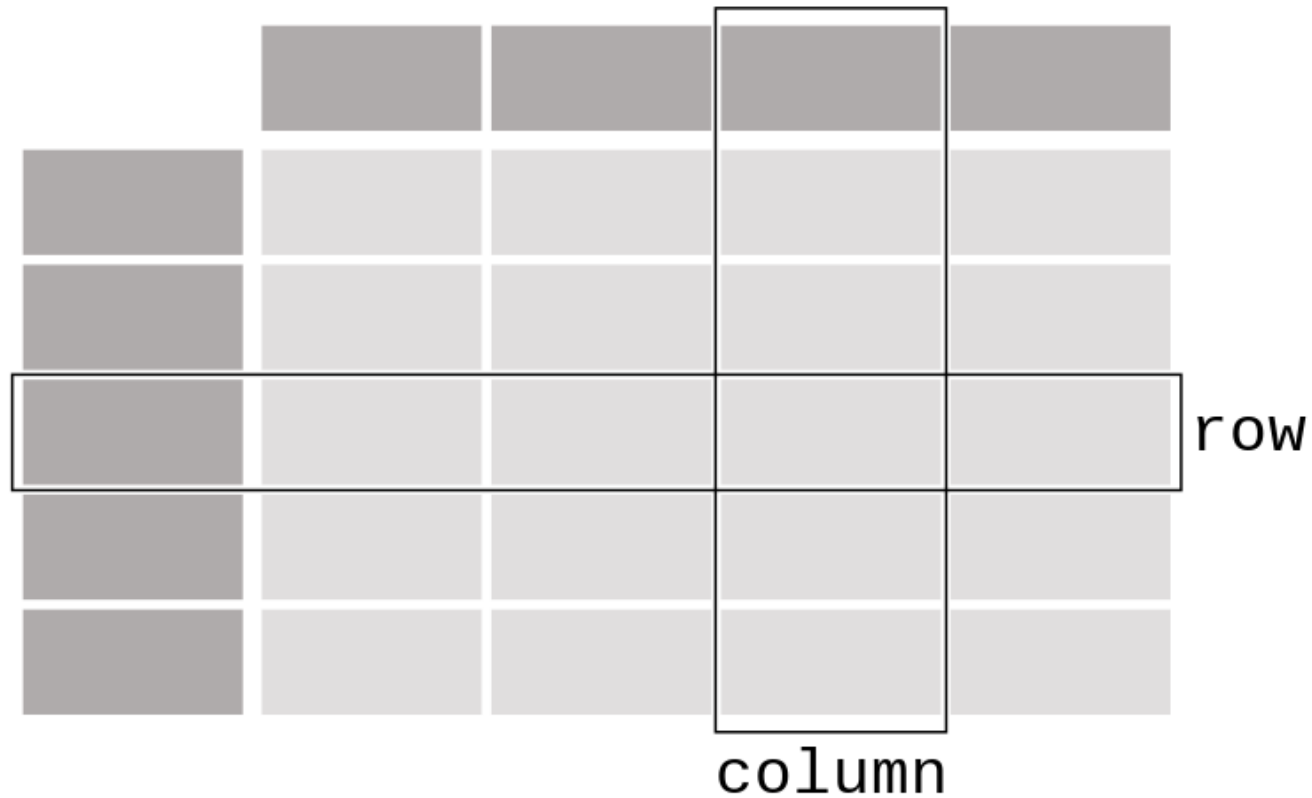
- **Primi strumenti moderna Data Analysis/Analytics**
 - Tipologie di analisi
 - Descrittiva (→ Analysis)
 - Predittiva/Prescrittiva (→ Analytics)
 - Preparazione dei dati (“Data Wrangling”)
 - Visualizzazione dei dati per primi rapporti statistici
 - “Affaccio” sul Machine Learning
- **L'uso a livello iniziale di alcuni strumenti di lavoro**
 - Python-based (pandas, Matplotlib, seaborn, Plotly)
 - KNIME Analytics Platform (?)

Dettagli

- **Durata**
 - 20 ore
- **Prerequisiti**
 - (Obbligatori) Curiosità, apertura mentale, attitudine alla novità
 - (Preferibile) Familiarità di base con:
 - Linguaggi di programmazione (es. Python, R, Java)
 - Installazione applicazioni (es. IDE, Docker) su comuni sistemi operativi (Windows, macOS, Linux)
- **Modalità**
 - (Un po' di) Teoria
 - Esempi ed esercizi (in aula e a casa)
 - Test finale (quiz da 10 domande + mini lab)

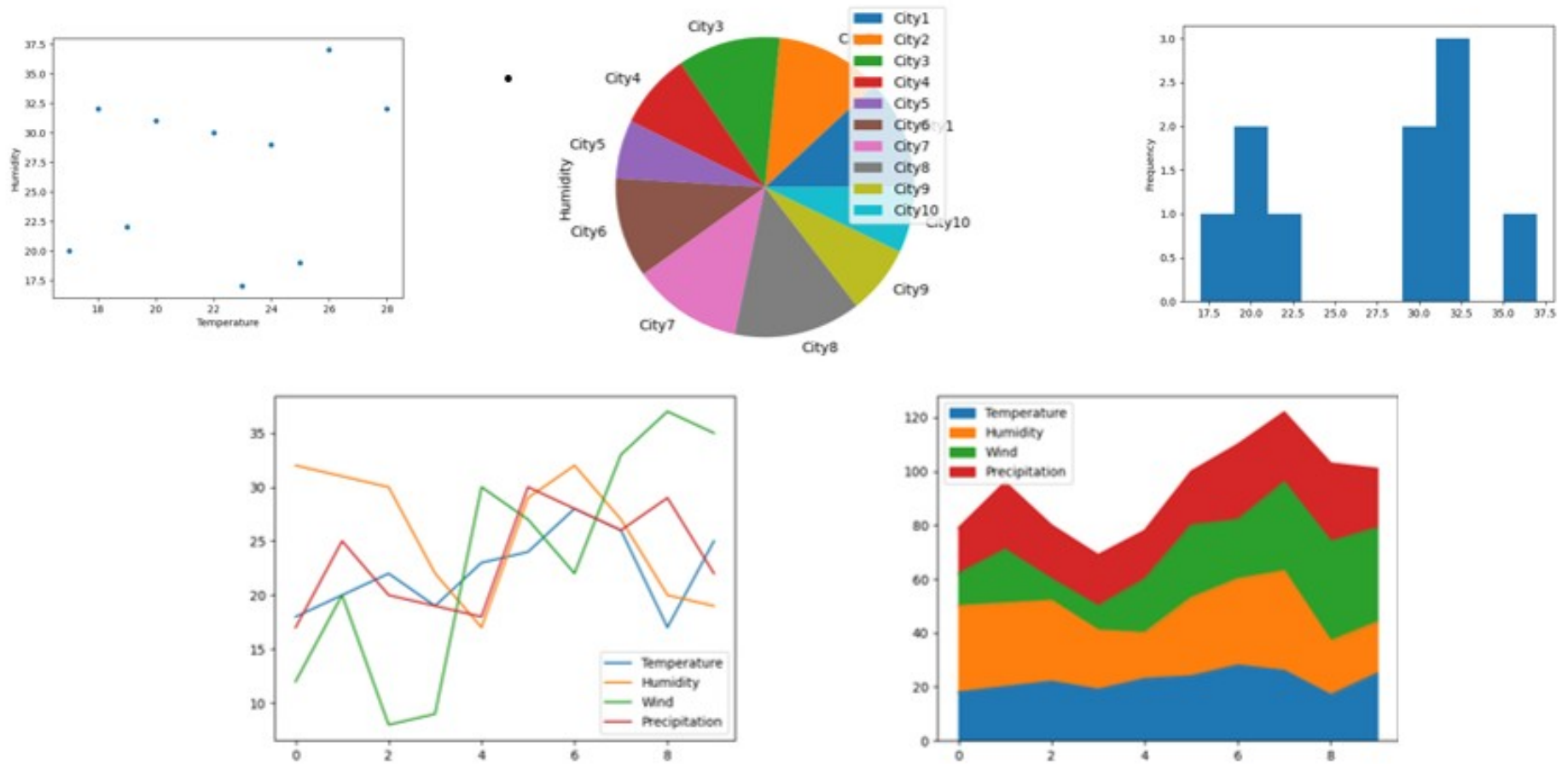
Domande?

DataFrame



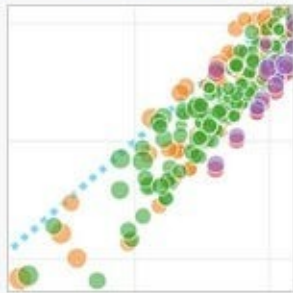
Python-based (pandas)



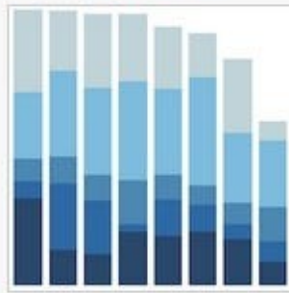


Python-based (Matplotlib)
matplotlib

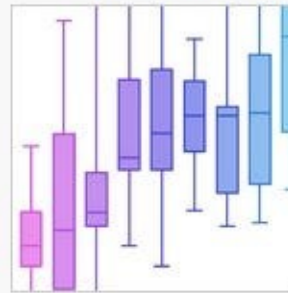
Line and Scatter Plots



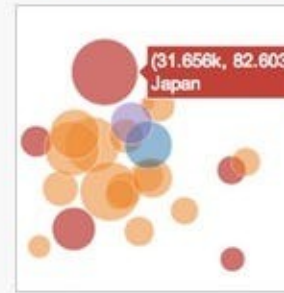
Bar Charts



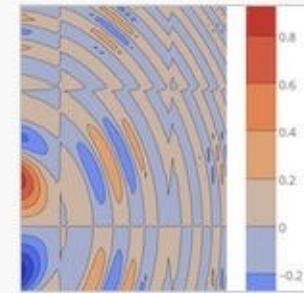
Box Plots



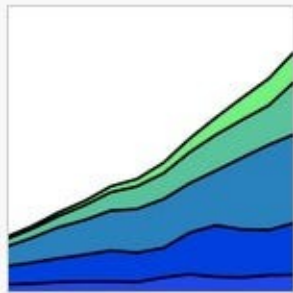
Bubble Charts



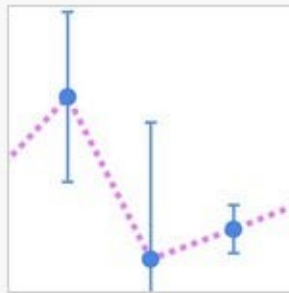
Contour Plots



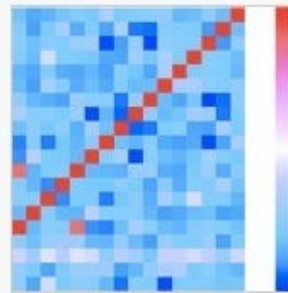
Filled Area Plots



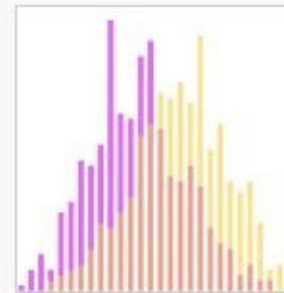
Error Bars



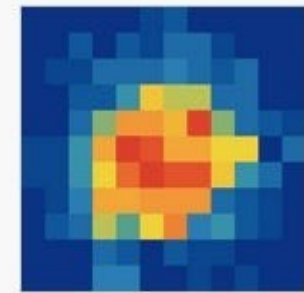
Heatmaps



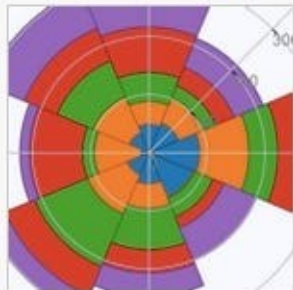
Histograms



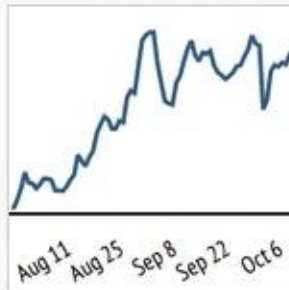
2D Histograms



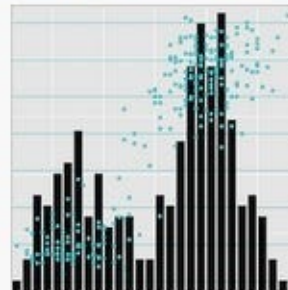
Polar Charts



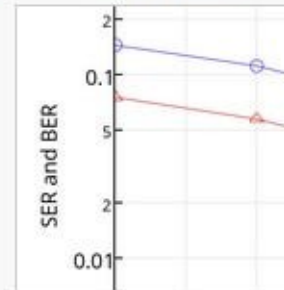
Time Series



Multiple Chart Types



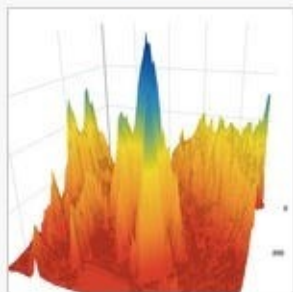
Log Plots



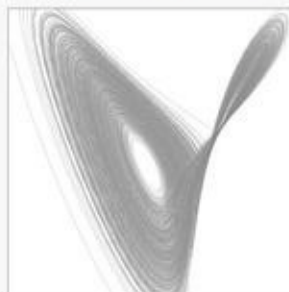
3D Scatter Plots



3D Surface Plots



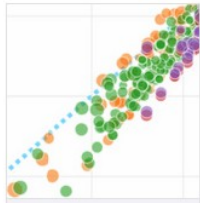
3D Line Plots



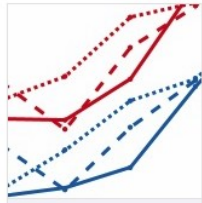
Python-based (seaborn)



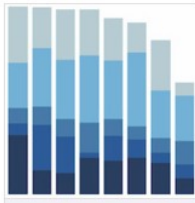
Basic Charts



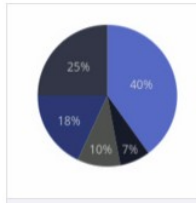
Scatter Plots



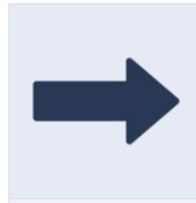
Line Charts



Bar Charts

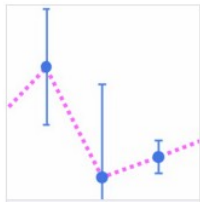


Pie Charts

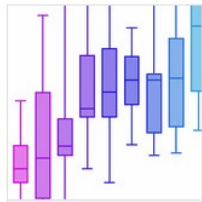


More Basic Charts

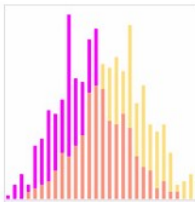
Statistical Charts



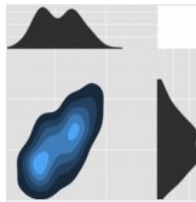
Error Bars



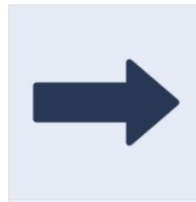
Box Plots



Histograms

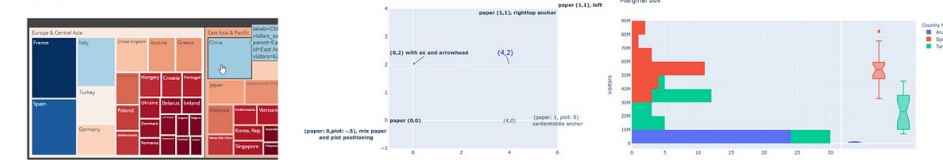
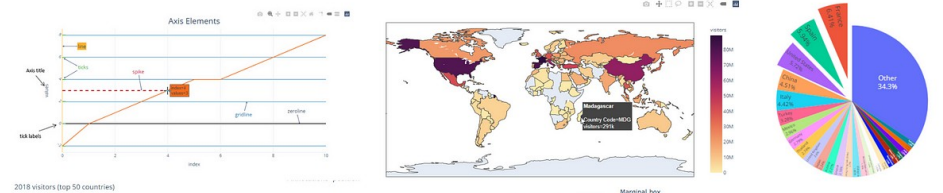


2d Density Plots



More Statistical Charts

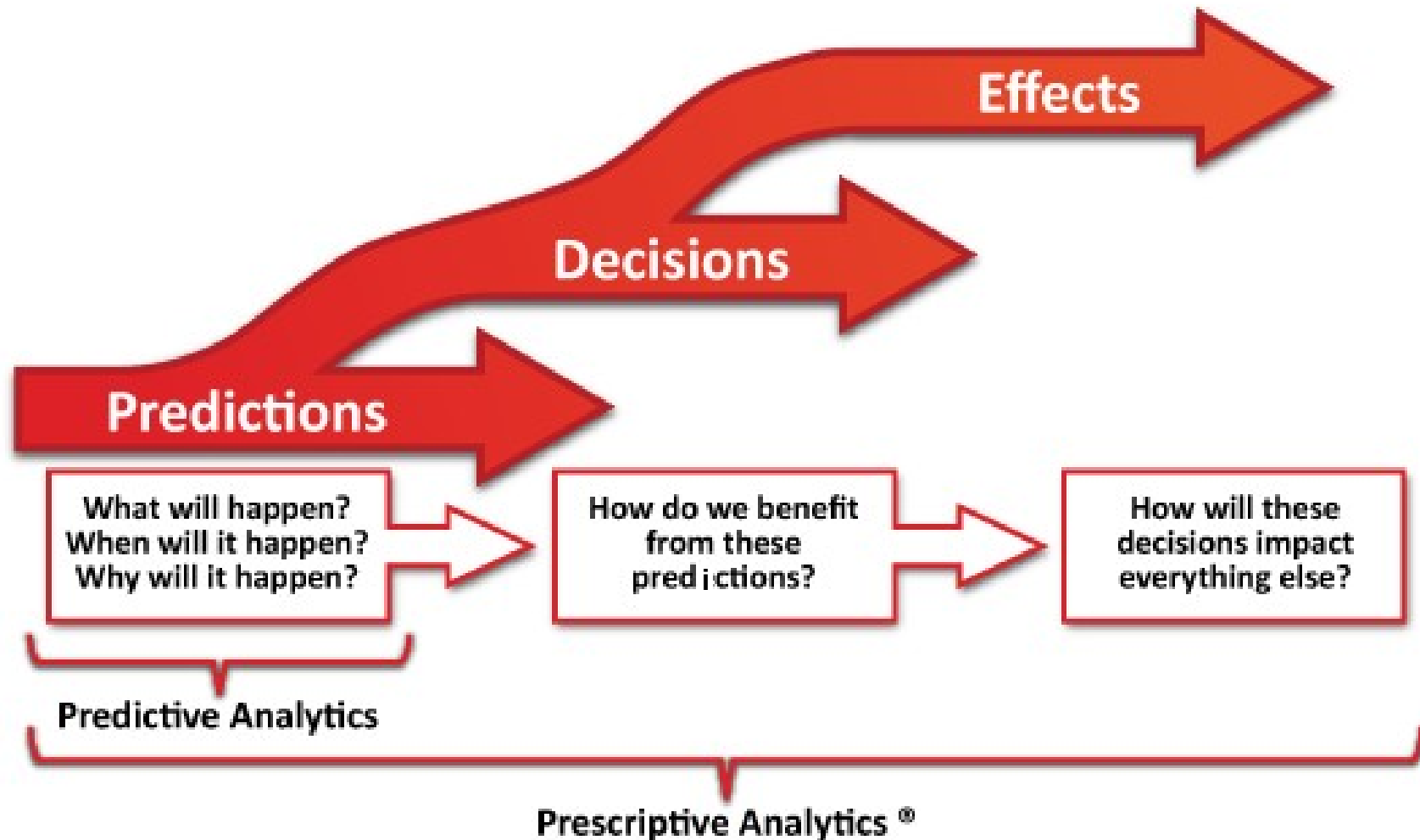
Python-based (Plotly)



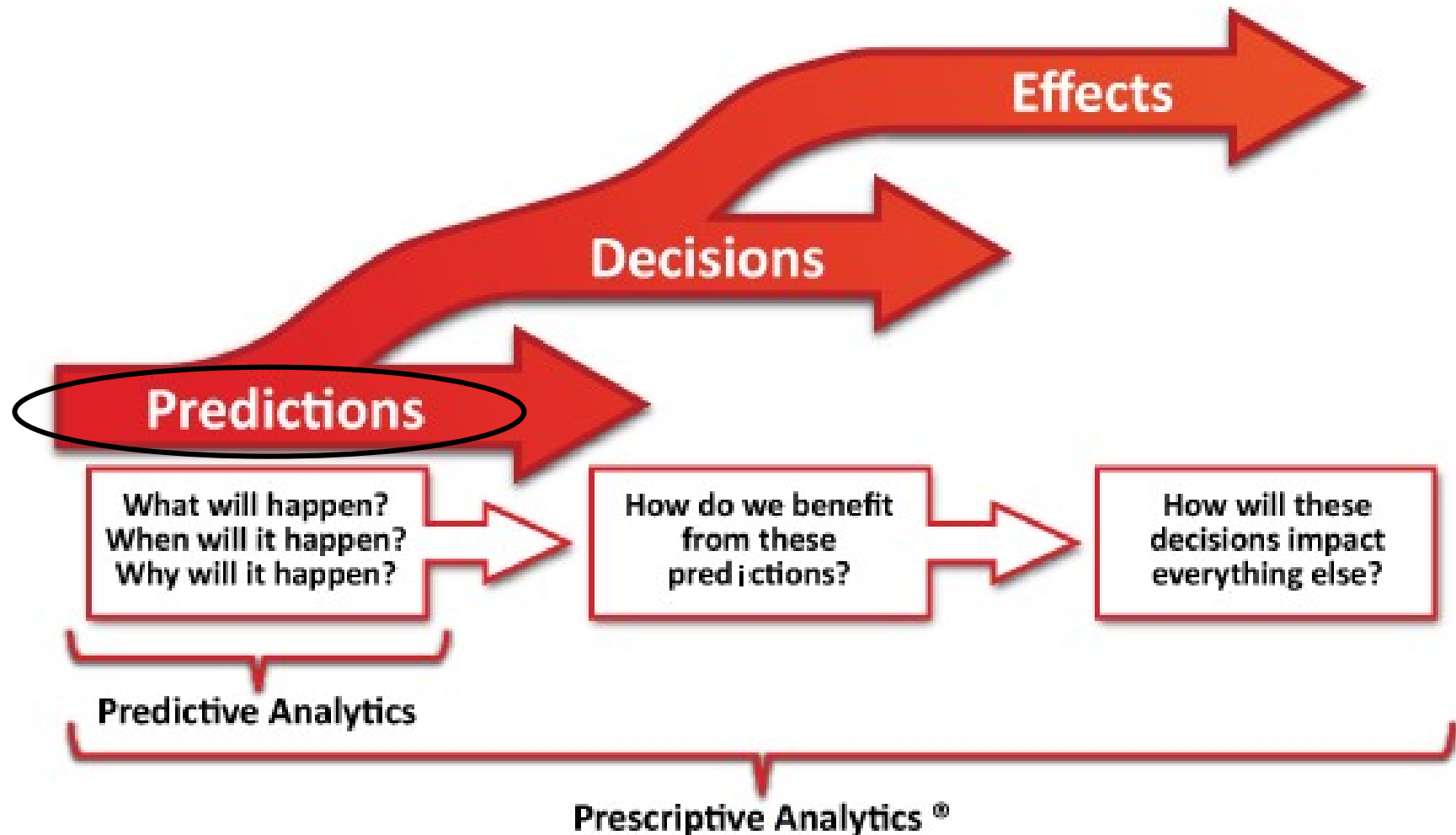
Data analysis vs. Data analytics

- Data analysis
 - Dati strutturati (modello relazionale)
 - Tecnologie SQL (RDBMS, BI, DWH)
 - Focus (principalmente) sul passato
 - “Human-driven evidence” (es. insiemi di regole)
- Data analytics
 - Dati strutturati e non strutturati (es. data lakes)
 - Tecnologie SQL e NoSQL
 - Focus (principalmente) sul futuro
 - “Machine-driven evidence” (es. modelli intelligenti automatici)

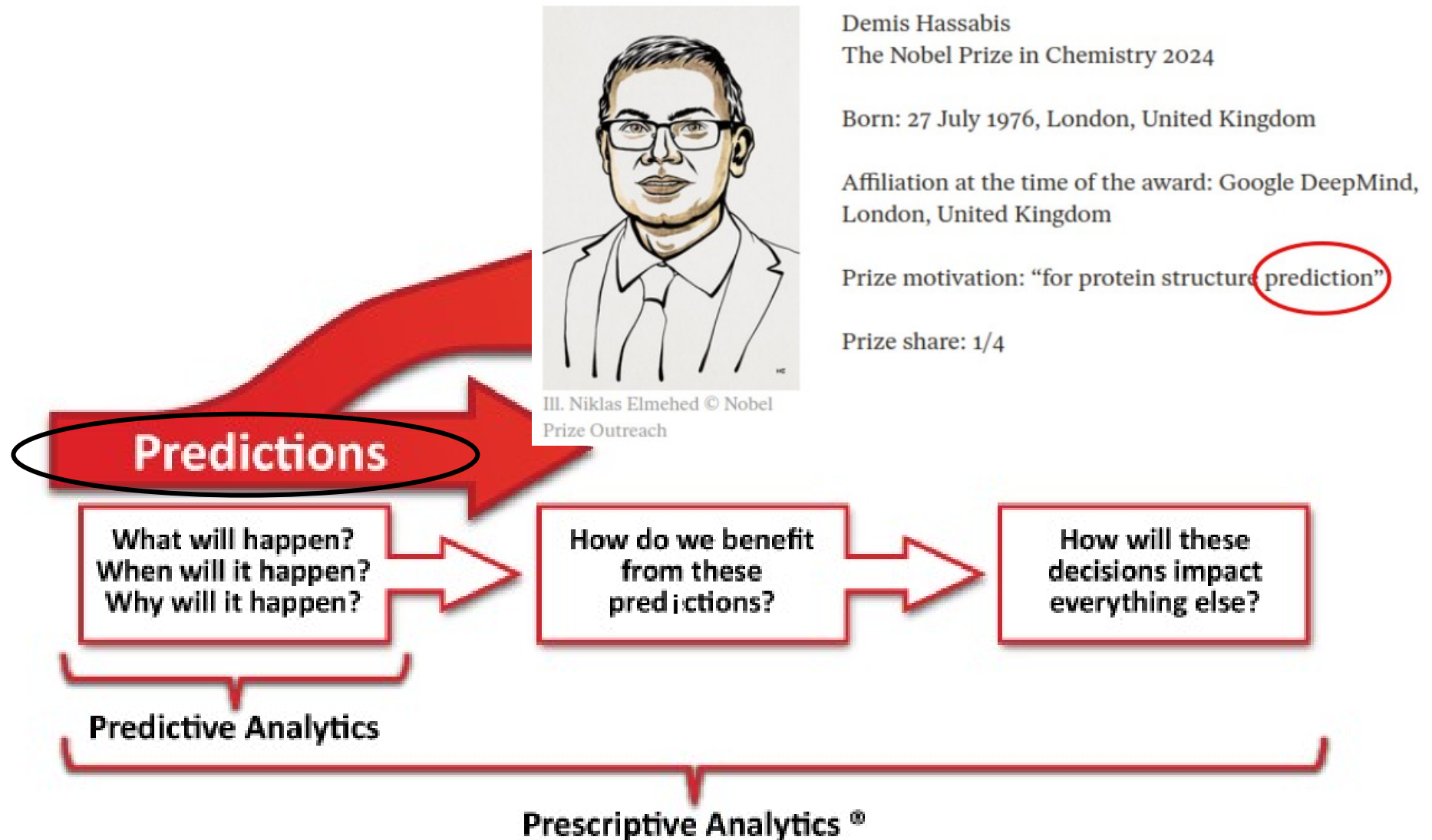
Data analytics...e il futuro!



Data analytics...e il futuro!



Data analytics...e il futuro!



Data analysis vs. Data analytics

- Data analysis (passato)
 - Descrittiva
 - Visualizzazioni
 - Report
 - Dashboard interattive
- Data analytics (futuro)
 - Predittiva
 - Diagnostica
 - Segmentazioni
 - Previsioni
 - Prescrittiva
 - Scelte/azioni ottimizzate
 - Modelli di propensione
 - Simulatori
 - Agenti generativi

Data analysis vs. Data analytics

- Data analysis (passato)

- Descrittiva

- Visualizzazioni
 - Report
 - Dashboard interattive

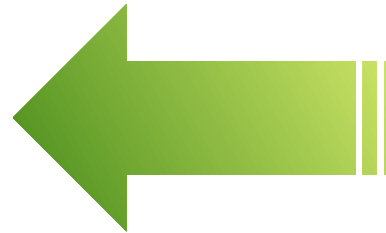
- Data analytics (futuro)

- Predittiva

- Diagnostica
 - Segmentazioni
 - Previsioni

- Prescrittiva

- Scelte/azioni ottimizzate
 - Modelli di propensione
 - Simulatori
 - Agenti generativi



Unità IA.1

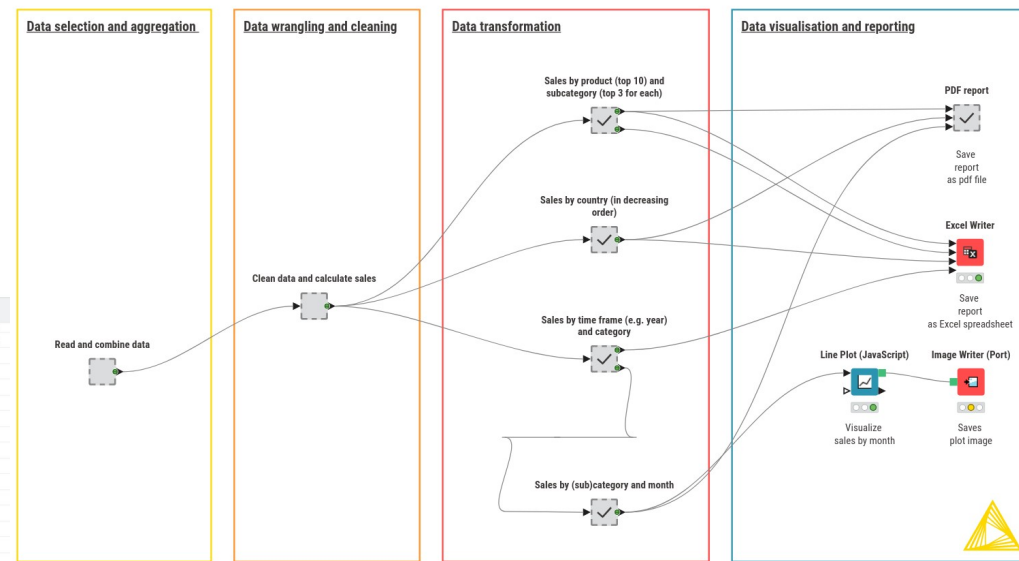
Esempi di analisi

Provate ad indovinare il tipo di analisi in ogni esempio

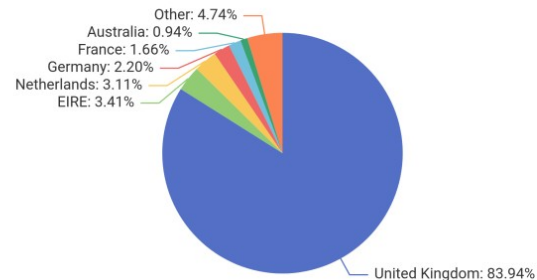
KNIME AP (Manipolazione)

Rows: 675802 | Columns: 10

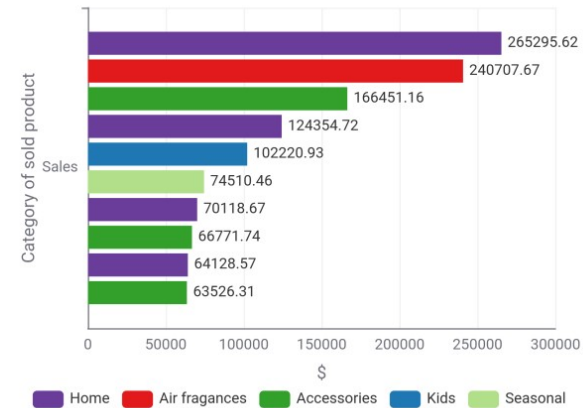
#	RowID	Invoice String	StockCode	Quantity Number (integer)	Price Number (double)	Customer_ID String	Invoice_time String	Country String	Description String	Category String	Subcategory String
47415	Row47	554789	22561	48	1.65	16966	D26/5/19T12:07:00	United Kingdom	WOODEN SCHOOL COLO	Stationery	Desktop Stationery
47415	Row47	554789	21703	120	0.42	16966	D26/5/19T12:07:00	United Kingdom	BAG 125g SWIRLY MARBL	Accessories	Bags
47415	Row47	554789	22560	48	1.06	16966	D26/5/19T12:07:00	United Kingdom	TRADITIONAL MODELLIN	Hobbies	Crafts
47415	Row47	C554790	22561	-4	1.45	14911	D26/5/19T12:10:00	EIRE	JAM MAKING SET PRINT	Home	Kitchen
47415	Row47	554791	22561	4	1.45	14911	D26/5/19T12:12:00	EIRE	JAM MAKING SET PRINT	Home	Kitchen
47415	Row47	554792	21080	24	0.85	12547	D26/5/19T12:38:00	Spain	SET/20 RED RETROSPOT	Home	Paper
47416	Row47	554792	POST	1	28	12547	D26/5/19T12:38:00	Spain	POSTAGE	Others	Code
47416	Row47	C554793	84077	-96	0.11	12901	D26/5/19T12:41:00	United Kingdom	WORLD WAR 2 GLIDERS A	Kids	Toys
47416	Row47	554794	23209	10	1.65	12720	D26/5/19T12:41:00	Germany	LUNCH BAG DOILEY PAT	Accessories	Bags
47416	Row47	554794	23203	10	2.08	12720	D26/5/19T12:41:00	Germany	JUMBO BAG VINTAGE DO	Accessories	Bags
47416	Row47	554794	22963	12	0.85	12720	D26/5/19T12:41:00	Germany	JAM JAR WITH GREEN LI	Home	Kitchen
47416	Row47	554794	22962	12	0.85	12720	D26/5/19T12:41:00	Germany	JAM JAR WITH PINK LI	Home	Kitchen
47416	Row47	554794	22961	12	1.45	12720	D26/5/19T12:41:00	Germany	JAM MAKING SET PRINT	Home	Kitchen
47416	Row47	554794	23111	2	12.5	12720	D26/5/19T12:41:00	Germany	PARISIENNE SEWING BOX	Hobbies	Crafts
47416	Row47	554794	23112	2	7.5	12720	D26/5/19T12:41:00	Germany	PARISIENNE CURIO CAB	Home	Furniture



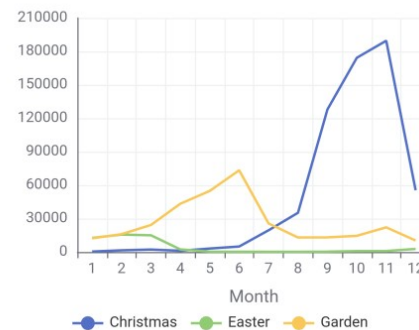
Top sales by country (%)



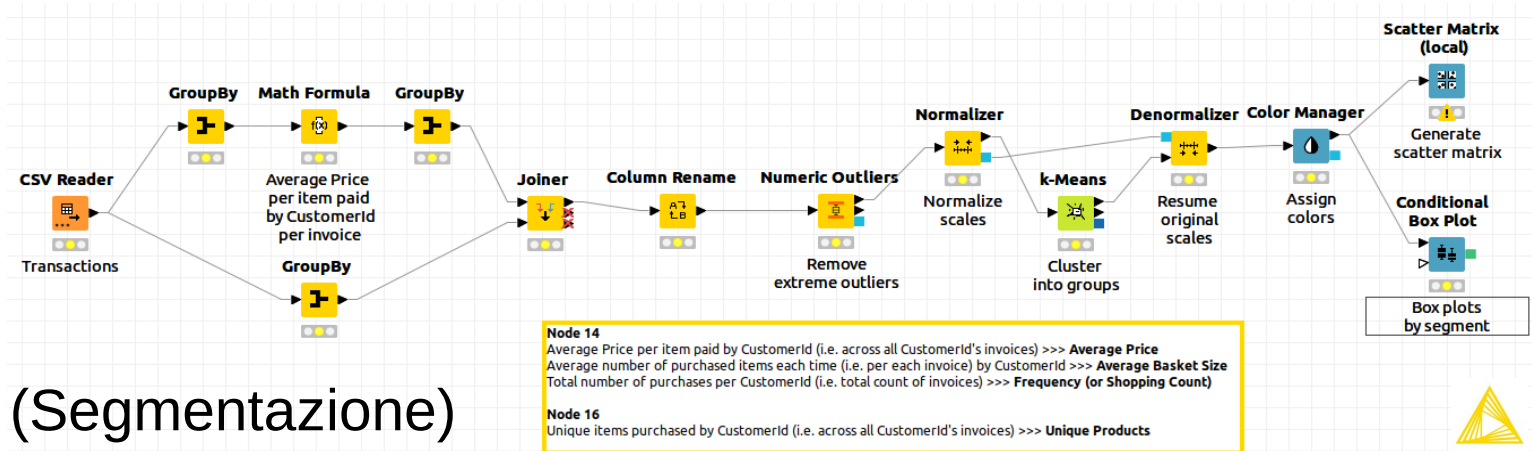
Top 10 sales by product (in \$)



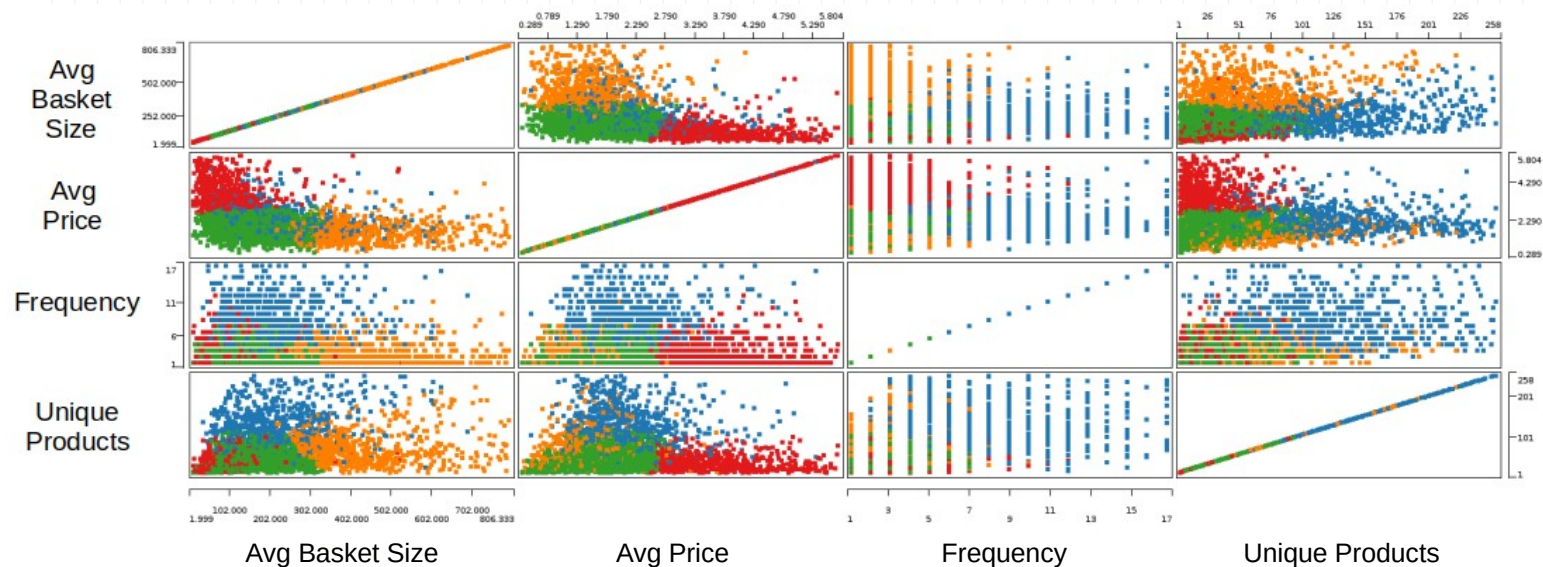
Sales of some items across the year (in \$)



Apriamo un report pdf...



KNIME AP (Segmentazione)



- **Segmento/cluster blu:** ~700 clienti, alta “Frequenza acquisto”, alta “Varietà prodotti” (**ABITUALI, CURIOSI, APERTI**)*
- **Segmento/cluster arancione:** ~600 clienti, alta “Dimensione carrello”, bassa “Frequenza acquisto”, basso “Prezzo medio”, bassa “Varietà prodotti” (**ABITUALI, RISPARMIATORI, RIPETITIVI**)°
- **Segmento/cluster verde:** ~1900 clienti, bassa “Frequenza acquisto”, basso “tutto” (**OCCASIONALI**)^
- **Segmento/cluster rosso:** ~600 customers, bassa “Dimensione carrello”, bassa “Varietà prodotti”, bassa “Frequenza acquisto”, alto “Prezzo medio” (**TOP, PREMIUM, LUSSO**)#

* Es. Invio frequente di comunicazioni, buoni sconto, campioni gratuiti

° Es. Segnalare novità di assortimento prodotti d'interesse, offerte speciali per grosse spese

^ Es. Sconti su prodotti già acquistati, informazione su simili/altre categorie merceologiche

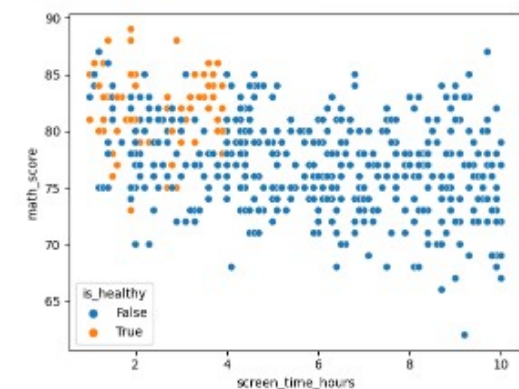
Es. Enfasi su qualità e valore, segnalazione assortimenti di “alta gamma”

Demo

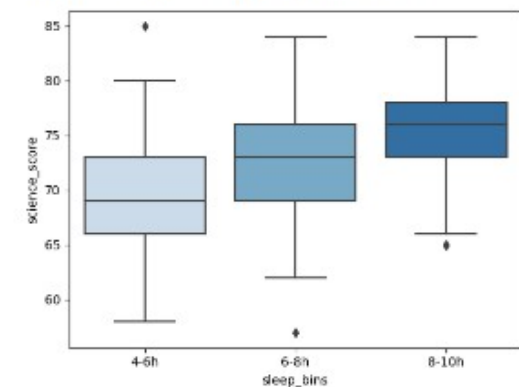
Una piccola dashboard con Python e Docker per cominciare a esplorare...

Some Habits Dashboard

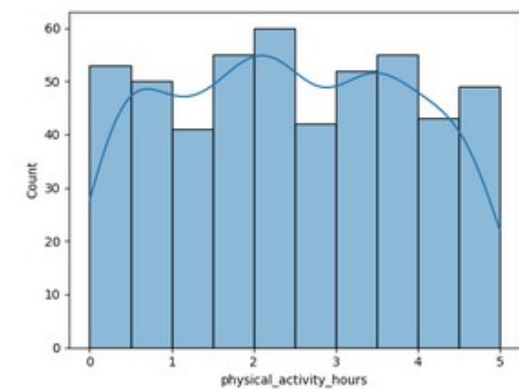
Screen Time vs. Math Score



Sleep Hours (Binned) vs. Science Score

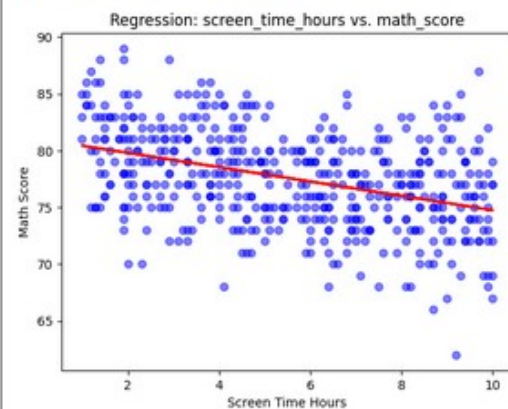


Physical Activity Distribution



Regression Result

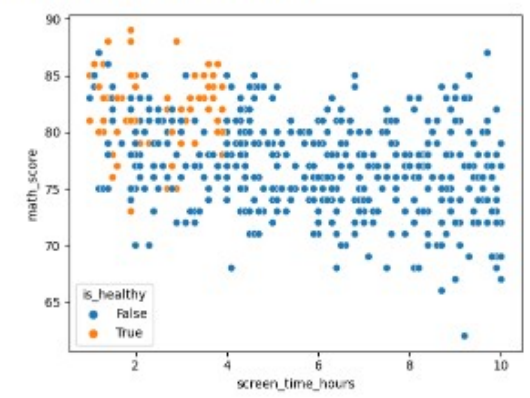
Impact of 1h screen_time_hours on math_score: -0.63 points



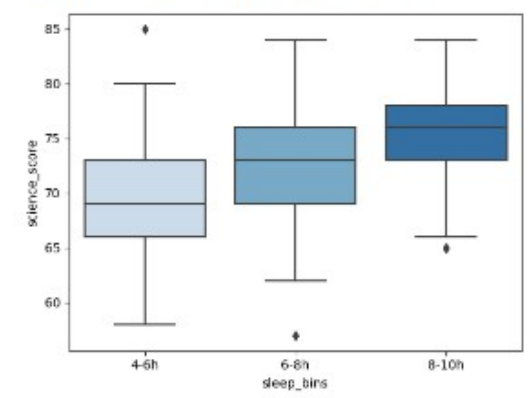
[Back](#)

Some Habits Dashboard

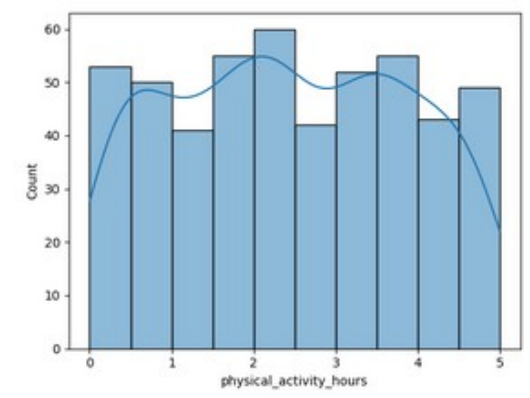
Screen Time vs. Math Score



Sleep Hours (Binned) vs. Science Score

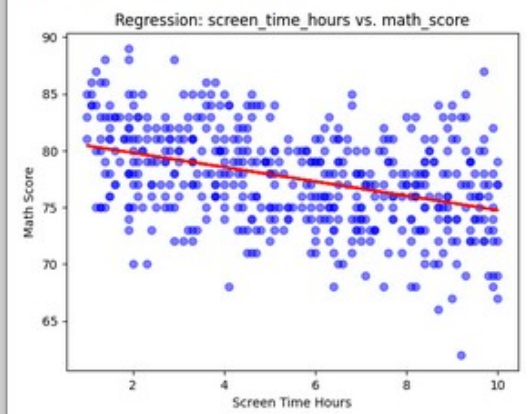


Physical Activity Distribution



Regression Result

Impact of 1h screen_time_hours on math_score: -0.63 points



[Back](#)

```
def plot_regression(x_col, y_col):
    plt.figure()
    X = df[[x_col]]
    y = df[y_col]
    model = LinearRegression().fit(X, y)
    plt.scatter(X, y, color='blue', alpha=0.5)
    plt.plot(X, model.predict(X), color='red', linewidth=2)
    plt.xlabel(x_col.replace("_", " ").title())
    plt.ylabel(y_col.replace("_", " ").title())
    plt.title(f"Regression: {x_col} vs. {y_col}")
    buf = io.BytesIO()
    plt.savefig(buf, format='png', bbox_inches='tight')
    plt.close()
    return base64.b64encode(buf.getvalue()).decode('utf-8'), model.coef_[0]
```

```
plots = [
    ("Screen Time vs. Math Score",
     lambda: sns.scatterplot(data=df, x="screen_time_hours", y="math_score", hue="is_healthy")),

    ("Sleep Hours (Binned) vs. Science Score",
     lambda: sns.boxplot(data=df, x="sleep_bins", y="science_score", palette="Blues")),

    ("Physical Activity Distribution",
     lambda: sns.histplot(data=df, x="physical_activity_hours", kde=True, bins=10))
]
```

Questa lezione

- Introduzione alla libreria **pandas**
- Concetto di DataFrame (“contenitore” di dati)
- Propedeutico ad analisi e visualizzazioni
- Primi esempi ed esercizi

Questa lezione

- Introduzione alla libreria **pandas**
- Concetto di DataFrame (“contenitore” di dati)
- Propedeutico ad analisi e visualizzazioni
- Primi esempi ed esercizi



[Apriamo il tutorial ufficiale...](#)