

Red Hat
Summit

Connect

AI4GRINS:

un progetto per rivoluzionare la Data Visualization
attraverso l'Intelligenza Artificiale Generativa

Massimiliano Bellifemine, Head of Digital Factory Cloud & Data Driven
Applications @ Exprivia

Riccardo Toti, Head of Artificial Intelligence @ Seacom Srl

Roma, 07/11/24



Agenda

- Che cosa è **GRINS**
- Organizzazione di **GRINS**
- Che cos'è **AMELIA**
- Bandi a cascata
- **AI4GRINS**
- Come l'AI sarà utilizzata
- MockUp implementativo
- Q&A

Che cosa è GRINS

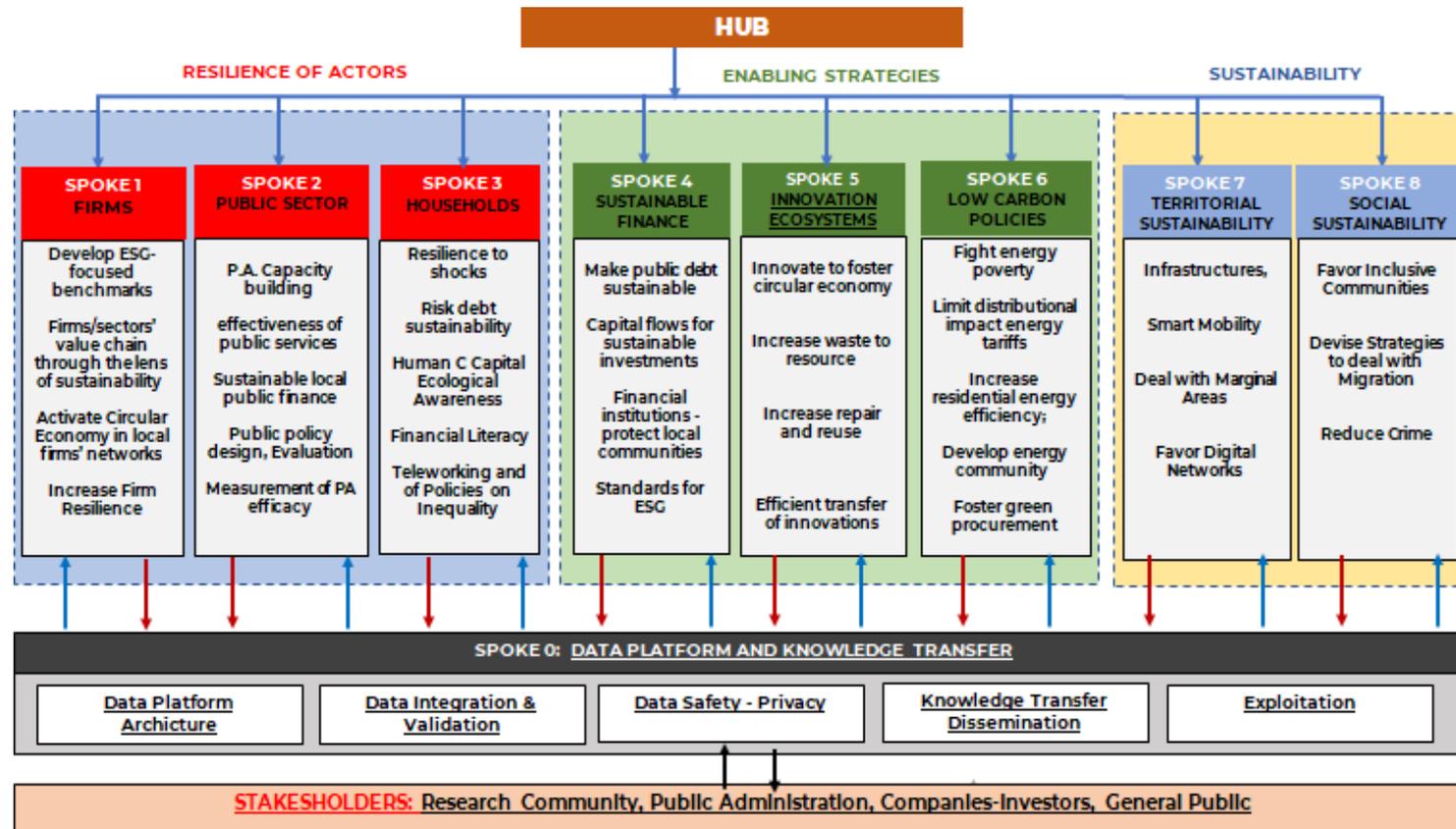
Growing, Resilient, Inclusive, Sustainable

- rientra tra i 14 accordi di partenariati estesi a università, centri di ricerca e aziende su tutto il territorio nazionale, selezionati dal MUR in seno al Piano Nazionale di Ripresa e Resilienza (PNRR).
- coinvolge una vasta gamma di competenze e risorse multidisciplinari provenienti da diverse fonti, tra cui università pubbliche, imprese private e consulenti.
- rientra nella tematica della «Sostenibilità economico-finanziaria dei sistemi e dei territori.»
- L'ultimo decennio ha portato sconvolgimenti macroeconomici senza precedenti, come la pandemia, la crisi dell'Unione Europea, il calo demografico, la crisi del debito sovrano, la migrazione e la transizione verde, impattando profondamente sulla vita quotidiana di tutti noi. Questi eventi hanno messo sotto pressione sia gli attori economici che i policy maker, lasciandoli impreparati.
- La ricerca deve agire tempestivamente, considerando questi problemi in una prospettiva sistemica e utilizzando i dati per favorire una crescita più resiliente, inclusiva e sostenibile, per imprese, famiglie e pubbliche amministrazioni.



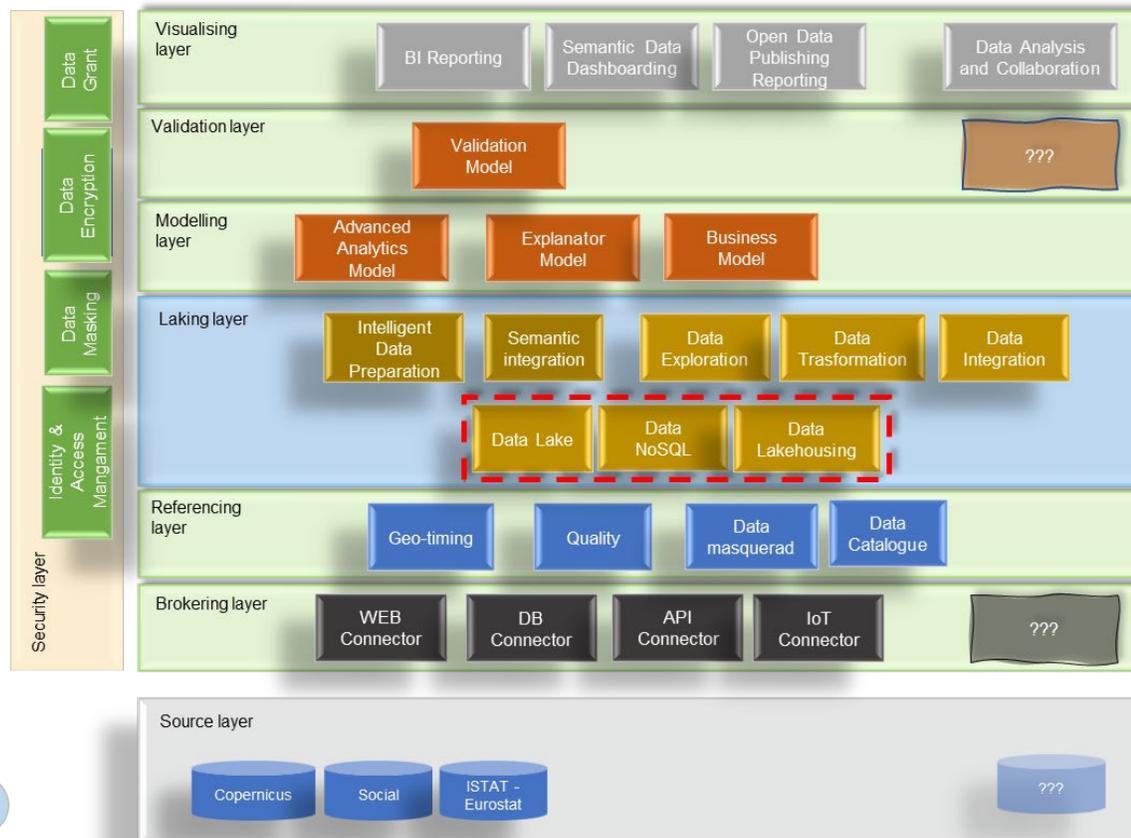
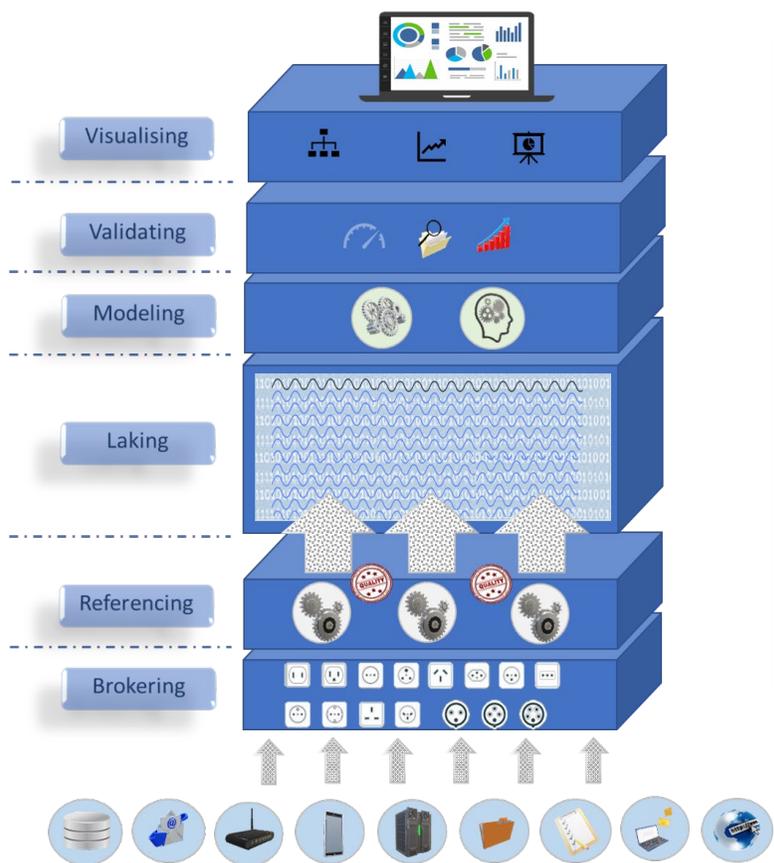
L'organizzazione

GRINS opera tramite 9 spoke con obiettivi differenti



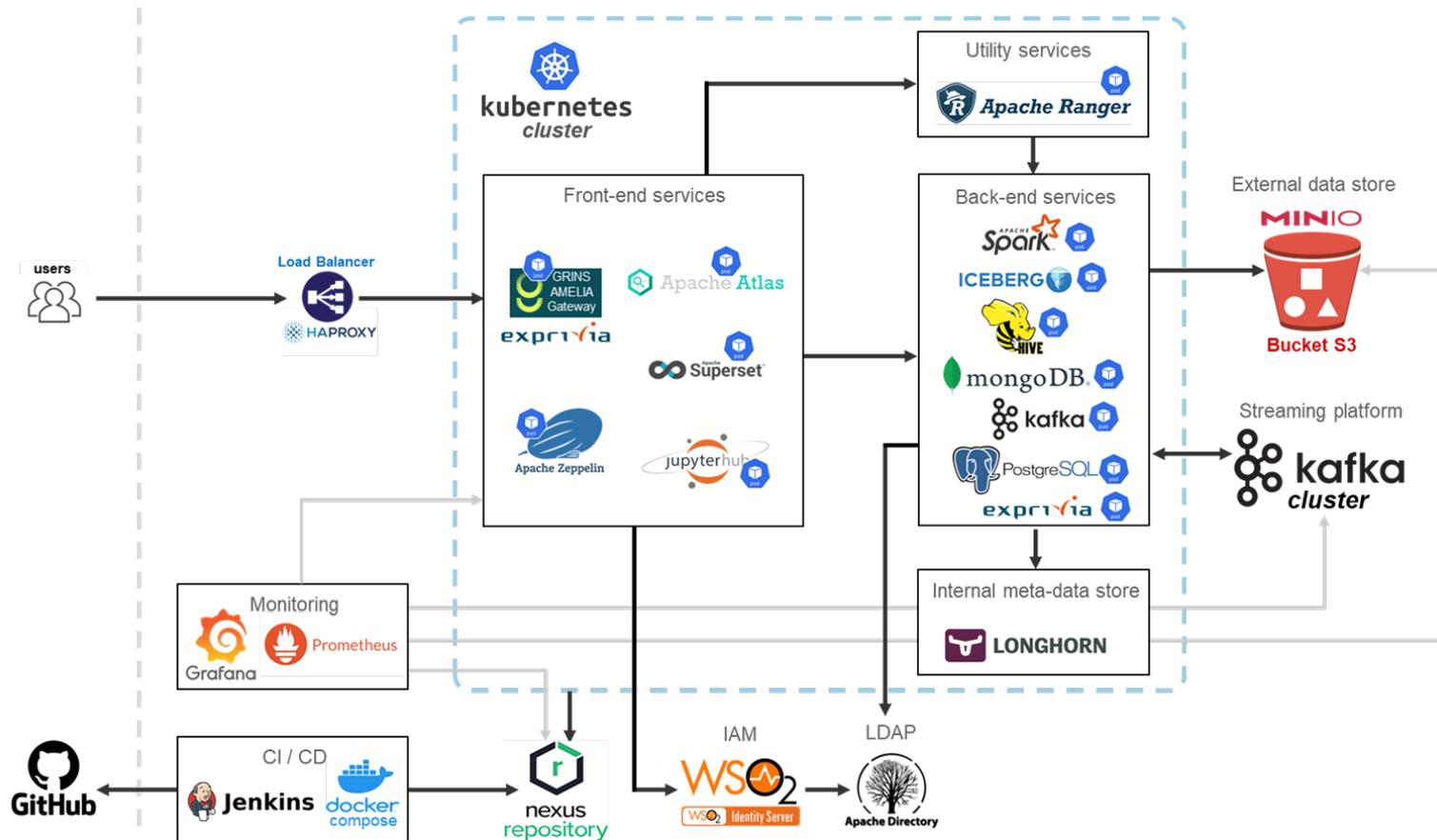
Che cosa è AMELIA

Una data platform a servizio della ricerca



Che cosa è AMELIA

Una data platform a servizio della ricerca



AMELIA e bandi a cascata

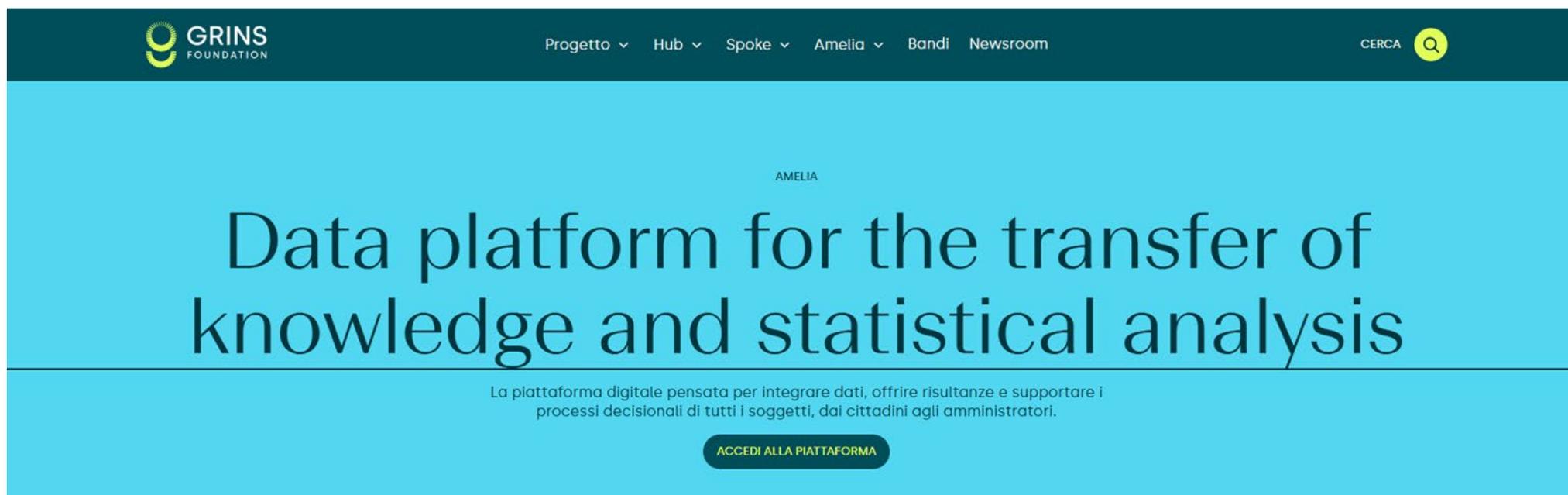
Ampliamento e attualizzazione

- ▶ Scopo: aggiungere servizi industriali alla piattaforma
- ▶ Rappresentazione grafica dei layer georeferenziati dei dati
- ▶ Deep learning per la produzione di dati sintetici
- ▶ Modelli di AI capaci di fare data extraction da documenti
- ▶ Interfacce con linguaggio naturale per interrogare i dati o produrre grafici automatici



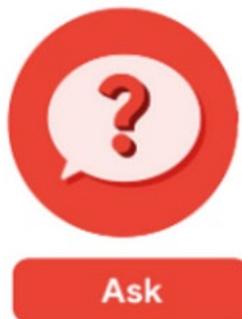
AI4GRINS

Il progetto di **Seacom** per il bando Grins per l'implementazione dell'**Intelligenza Artificiale** all'interno della piattaforma **Amelia**



The screenshot shows the Amelia platform website. The top navigation bar is dark teal with the GRINS FOUNDATION logo on the left and a search bar on the right. The main content area has a light blue background with the title 'Data platform for the transfer of knowledge and statistical analysis' in large, dark teal font. Below the title is a short description in Italian and a button labeled 'ACCEDI ALLA PIATTAFORMA'.

Business Questions



Ask questions and define the problem.

Utilizzare i LLM per suggerire dei **goal** e degli **insight automatici** per esplorare e analizzare più velocemente i propri dataset.

Suggerirà inoltre la visualizzazione con delle infografiche inerenti ai nostri dati.

Data Process tramite il linguaggio naturale



Prepare

Prepare data by collecting and storing the information.



Process

Process data by cleaning and checking the information.

Utilizzare il **linguaggio naturale** per modificare e processare automaticamente i dati affinché creino correttamente le visualizzazioni secondo le esigenze personali.

- *Rimuovi tutti i record NaN dal dataset*
- *Formatta tutti i record USA con Stati Uniti*

Data Analysis



Analyze

Analyze data to find patterns, relationships, and trends.

Estrazione automatica di **analisi statistiche** come min, max, moda, mediana per le feature numeriche e la **topologia** per le feature categoriche.

Inoltre potrà essere eseguito vari analisi avanzate come il clustering (utilizzando vari algoritmi **DBSCAN, K-MEANS**, ecc).

- *Visualizza **in accordo con i dati** la correlazione tra la variabile x e y*
- *Visualizza i dati in clusters utilizzando un **k=4***

Riepilogo dei dati

cars.json Questo set di dati contiene informazioni su varie auto, tra cui nome, chilometri per gallone, numero di cilindri, cilindrata, cavalli, peso in libbre, accelerazione, anno di fabbricazione e origine.

^ riepilogo dei dati | cars.json

Nome

stringa | modello_auto
Il nome o il modello dell'auto.
Valori univoci: 311

Visualizza i campioni

Miglia_per_gallone

numero | chilometraggio
Numero di miglia che l'auto può percorrere con un gallone di carburante.
Valori univoci: 129
minimo: 9 , massimo: 46,6 | valore standard: 7,816

Visualizza i campioni

Cilindri

numero | numero
Il numero di cilindri nel motore dell'auto.
Valori univoci: 5
minimo: 3 , massimo: 8 | standard: 1.000

Visualizza i campioni

4 5 6

Cilindrata

numero | cilindrata
La cilindrata del motore dell'auto, misurata in pollici cubi.
Valori univoci: 83
minimo: 68 , massimo: 455 | standard: 104.922

Visualizza i campioni

Potenza

numero | potenza
La potenza dell'auto.
Valori univoci: 93
minimo: 46 , massimo: 230 | std: 38.769

Visualizza i campioni

Peso_in_libbre

numero | peso
Il peso dell'auto in libbre.
Valori univoci: 356
minimo: 1613 , massimo: 5140 | valore di riferimento: 847.000

Visualizza i campioni

Accelerazione

numero | velocità
L'accelerazione dell'auto, misurata in secondi, da 0 a 60 mph.
Valori univoci: 96
minimo: 8 , massimo: 24,8 | valore standard: 2,803

Visualizza i campioni

Anno

data | anno
L'anno di fabbricazione dell'auto.
Valori univoci: 12
minimo: 1970-01-01 , massimo: 1982-01-01

Visualizza i campioni

Origine

categoria | posizione
Il paese in cui è stata prodotta l'auto.
Valori univoci: 3

Visualizza i campioni

🔍 Goal Exploration

Un elenco di obiettivi di esplorazione dei dati generati automaticamente (ipotesi) in base al riepilogo dei dati sopra riportato.

^ Obiettivi (5)

1 | Qual è la distribuzione dei chilometri per gallone tra le diverse origini?

Grafico a barre con 'Origine' sull'asse x e 'Miglia_per_gallone' sull'asse y
Questa visualizzazione utilizza i campi "Origin" e "Miles_per_Gallon" per confrontare l'efficienza del carburante tra diversi paesi di origine. Ciò potrebbe rivelare se le auto di determinati paesi tendono a essere più efficienti nei consumi.

2 | Come è cambiata la potenza media nel corso degli anni?

Grafico a linee con "Anno" sull'asse x e "Potenza media" sull'asse y
Questa visualizzazione utilizza i campi "Anno" e "Potenza" per tracciare i cambiamenti nella potenza delle auto nel tempo. Ciò potrebbe rivelare le tendenze nella produzione di auto e nelle preferenze dei consumatori.

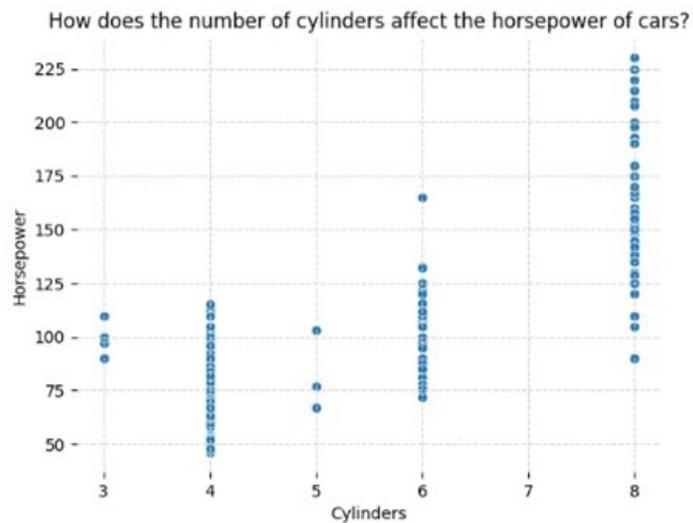
3 | Esiste una correlazione tra peso e chilometri per gallone?

Diagramma di dispersione con 'Weight_in_lbs' sull'asse x e 'Miles_per_Gallon' sull'asse y
Questa visualizzazione utilizza i campi 'Weight_in_lbs' e 'Miles_per_Gallon' per esplorare la relazione tra peso dell'auto ed efficienza del carburante. Ciò potrebbe rivelare se le auto più pesanti tendono a essere meno efficienti nei consumi.

4 | Qual è la distribuzione delle bombole nelle auto di diversa provenienza?

Grafico a barre impilate con 'Origine' sull'asse x e conteggio dei 'Cilindri' sull'asse y
Questa visualizzazione utilizza i campi "Origine" e "Cilindri" per confrontare i tipi di motore nei diversi paesi di origine. Ciò potrebbe rivelare se alcuni paesi tendono a produrre auto con un certo numero di cilindri.

Come il numero dei cilindri influenza i cavalli di un veicolo?



</> Python Code

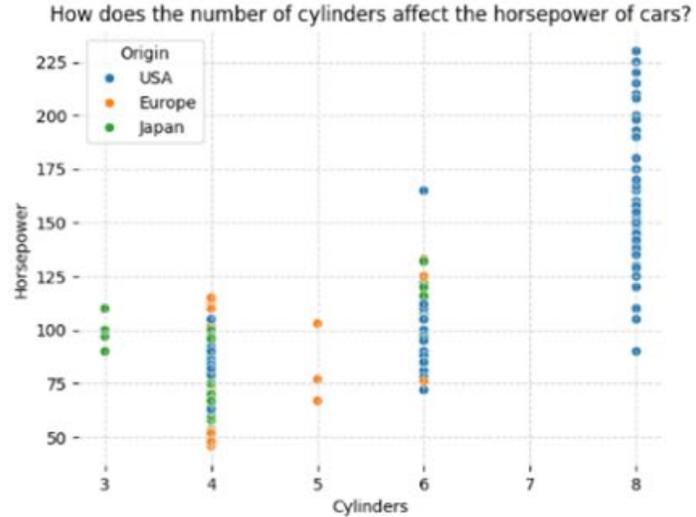
```
import pandas as pd
import matplotlib.pyplot as plt

# solution plan
# i. Create a scatter plot of Cylinders vs Horsepower

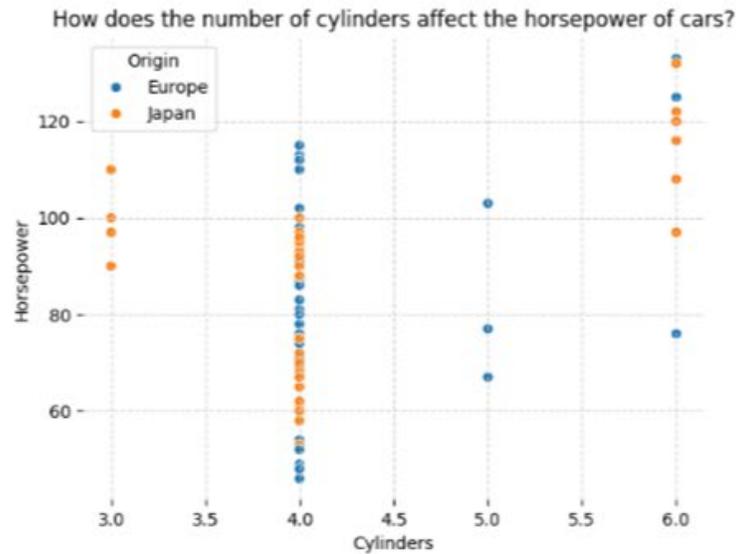
def plot(data: pd.DataFrame):
    sns.scatterplot(data=data, x='Cylinders', y='Horsepower')
    plt.title('How does the number of cylinders affect the horsepower of cars?', wrap=True)
    return plt

chart = plot(data)
```

Puoi modificare il grafico aggiungendo una distinzione in base all'originale delle auto?



Rimuovi tutte le auto provenienti dagli Stati Uniti?



</> Python Code

```
import pandas as pd
import matplotlib.pyplot as plt

# solution plan
# i. Create a scatter plot of Cylinders vs Horsepower

def plot(data: pd.DataFrame):
    sns.scatterplot(data=data[data['Origin'] != 'USA'], x='Cylinders', y='Horsepower', hue='Origin')
    plt.title('How does the number of cylinders affect the horsepower of cars?', wrap=True)
    return plt
```

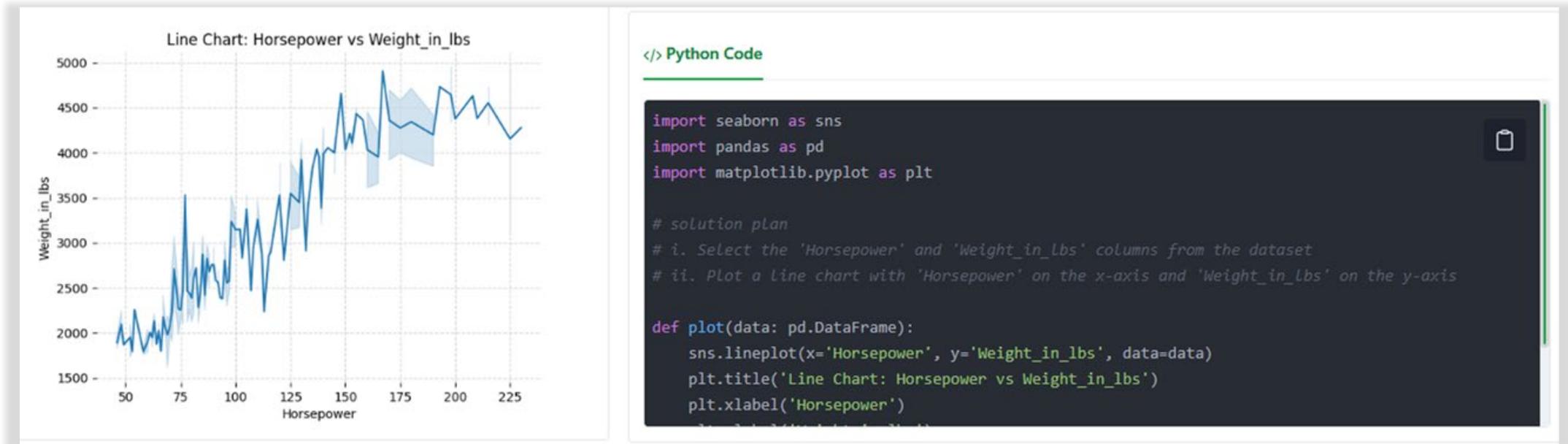


Visualization Generation

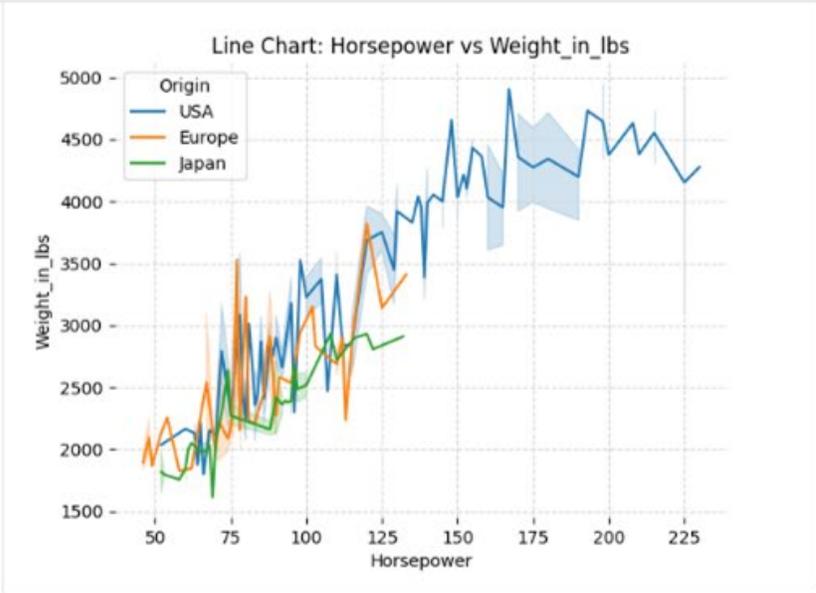
Select a goal above or describe a new visualization goal to generate a visualization.

line chart tra cavalli e peso delle auto?

 Generate



Aggiungi una diversificazione tra le auto provenienti da diversi paesi?



</> Python Code

```
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt

# solution plan
# i. Select the 'Horsepower' and 'Weight_in_lbs' columns from the dataset
# ii. Plot a line chart with 'Horsepower' on the x-axis and 'Weight_in_lbs' on the y-axis
# iii. Add a hue parameter to differentiate cars from different countries

def plot(data: pd.DataFrame):
    sns.lineplot(x='Horsepower', y='Weight_in_lbs', hue='Origin', data=data)
    plt.title('Line Chart: Horsepower vs Weight_in_lbs')
```

Red Hat
Summit

Connect

Q&A

Red Hat
Summit

Connect

Thank you

