Skills Required for Data Science Roles

Daniela Gonzalez
Ikonkar Khalsa
Matthew Gregorio

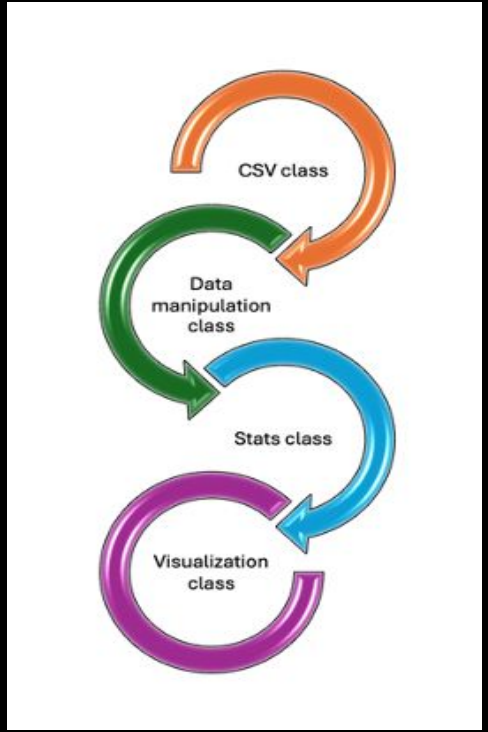
Goal

Discern the critical skills necessary for data science positions by developing a comprehensive set of modular classes.

1. Identify key skills for data science roles by analyzing the dataset.
2. Determine the most and least frequently mentioned skills in the dataset.
3. Uncover patterns suggesting proficiency in one skill may also require expertise in another for data science roles.

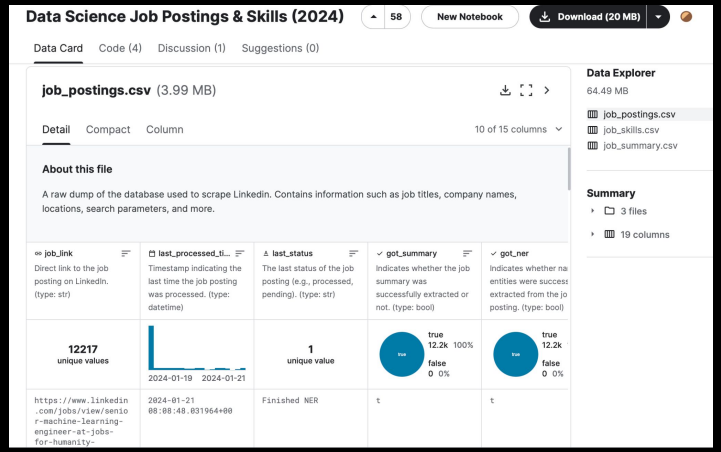
Objective

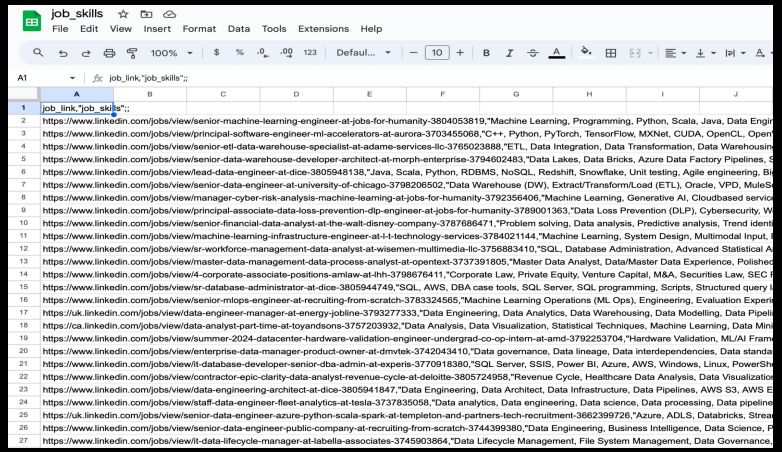
1. Create classes to import and open datasets
2. Design techniques for efficiently extracting pertinent information through data manipulation and management
3. Conduct data analysis to identify the 10 most and least required skills for data science roles and top 10 pairings
4. Utilize modules and classes to create visual representations that effectively conveys the results and correlations of the data



Data Set

- Kaggle
  - CSV File - Data Science Job Postings & Skills (2024)
- “This dataset provides a raw dump of data science-related job postings collected from Linkedin. It includes information about job titles, companies, locations, search parameters, and other relevant details.”





Classes & Methods

- CSV class
- def load_data(self):
  """
  Method load_data.
  Load data from a CSV file into self.data list.
  """
- def display_data(self, n=5):
  """
  Method display_data.
  Display the first n rows of loaded data.
  Args.: self and number of rows.
  """
- def tail_data(self, n=5):
  """
  Method: tail_data.
  Display the last n rows of loaded data.
  Args.: self and the number of rows to display
  """
- def info(self):
  """
  Method: info
  Display information about the loaded data.
  Args.: self.
  """

Data handling and exploration with CSV files (makes it easy for users to load, view, & understand data)

- Class Skill Counter
- class SkillCounter:
  """
  Class: SkillCounter. Designed to iterate over the dataframe
  and get information about the
  required skills for Data Science jobs.
  """
- def count_skills(self):
  """
  Method: count_skills.
  Count the frequency of individual skills and combinations of skills in the dataset.
  Returns:
  skill_counts (dict): Dictionary containing counts of individual skills.
  top_skills (list): List of tuples containing the top 10 most common individual skills.
  top_combinations (list): List of tuples containing the top 10 most common combinations of skills.
  """
- def least_required_skills(self):
  """
  Method: least_required_skills.
  Find the 10 least required skills in the dataset.
  Returns:
  least_required_skills (list): List of tuples containing the 10 least required skills and their counts.
  """

Counting frequencies that returns dictionaries with counts of individual skills and tuples.

Classes & Methods

```
class WordCloudGenerator:
    """
    Class: WordCloudGenerator. Designed to generate a word cloud from a list of the 10 most required skills.
    """

def generate_wordcloud(self):
    """
    Method: generate_wordcloud.
    Generate a word cloud from
    the top 10 most required skills.
    """
```

Generates word cloud visualization from the top 10 most required skills. Text data are visually represented by the size of the words which will indicate its frequency/importance.

```
class SkillsVisualizer:
    """
    Class: SkillsVisualizer
    Description: This class is designed to visualize skills combinations using a chord diagram.

    Attributes:
        top_combinations (list): List of tuples containing the top skill combinations and their counts.

    Methods:
        visualize_chord_diagram: Visualizes skill combinations using a chord diagram.
    """

    def visualize_chord_diagram(self):
        """
        Method: visualize_chord_diagram
        Visualizes skill combinations using a chord diagram.
        """
```

This method is responsible for visualizing skill combinations using a chord diagram. It shows the relationships between data points, in this case it the diagram shows the relationship between different skill combinations.

Results

Top 10 required skills:

[('Python', 4391), ('SQL', 4215), ('Communication', 2421), ('Tableau', 1624), ('AWS', 1615), ('Data Visualization', 1508), ('R', 1503), ('Data Analysis', 1479), ('Machine Learning', 1398), ('Spark', 1369)]

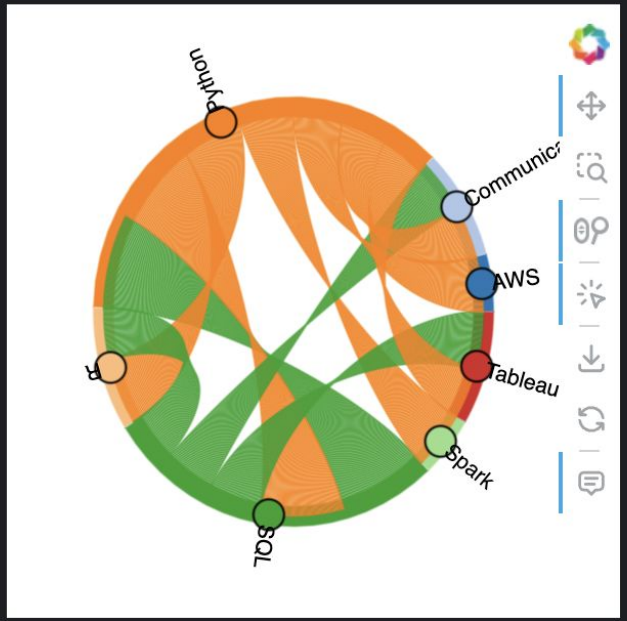
The 10 least required skills:

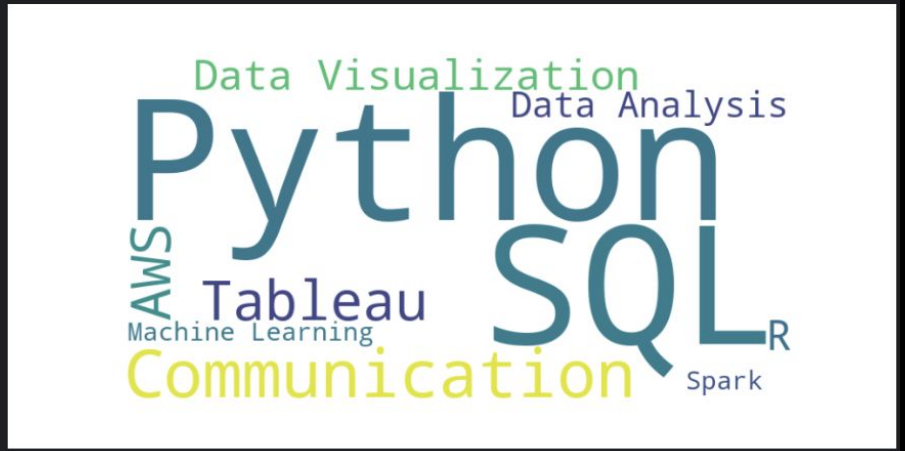
[('Multi node Database Paradigms', 1), ('SIMD programming models', 1), ('MLspecific accelerators', 1), ('Linux/unix environments', 1), ('Computer vision deep learning models', 1), ('ML software and hardware technology', 1), ('Inference on edge platforms', 1), ('Cloud ML training pipelines', 1), ('HPC experience', 1), ('Performance troubleshooting', 1)]

Top 10 most common skill combinations:

[(('SQL', 'Python'), 1407), (('Python', 'SQL'), 1332), (('Python', 'R'), 1005), (('Python', 'Spark'), 872), (('SQL', 'Tableau'), 866), (('SQL', 'Communication'), 818), (('Python', 'AWS'), 817), (('SQL', 'R'), 766), (('Python', 'Communication'), 751), (('Python', 'Tableau'), 743)]

Results


Python
Communica
AWS
Tableau
Spark
SQL
R


Data Visualization
Data Analysis
Python
SQL
AWS
Tableau
R
Machine Learning
Communication
Spark