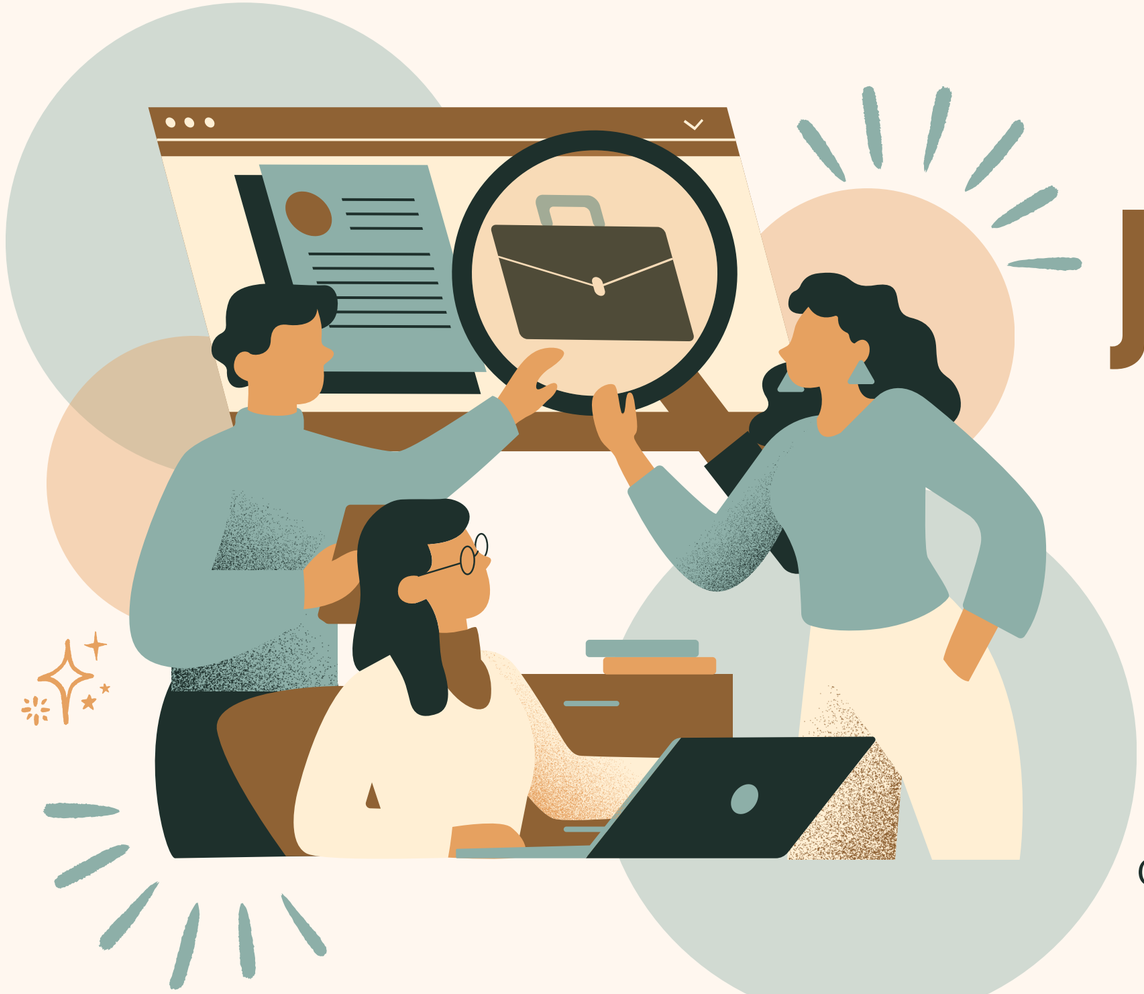
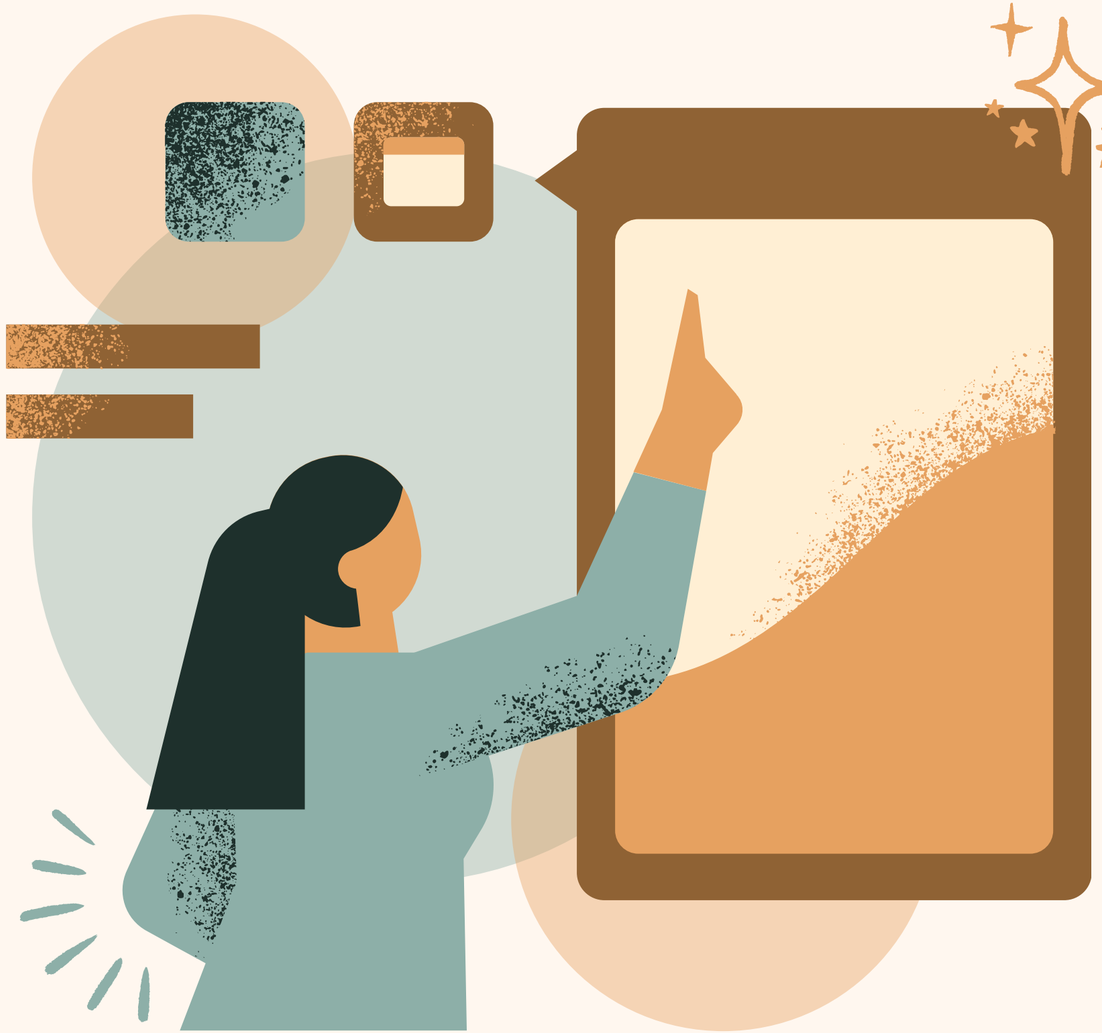


JOBS AND SALARIES IN

Data science

Group members: Zhanel Zhanbyrbayeva, Pelyno Olena, Yoshida Marie

TABLE OF *Contents*

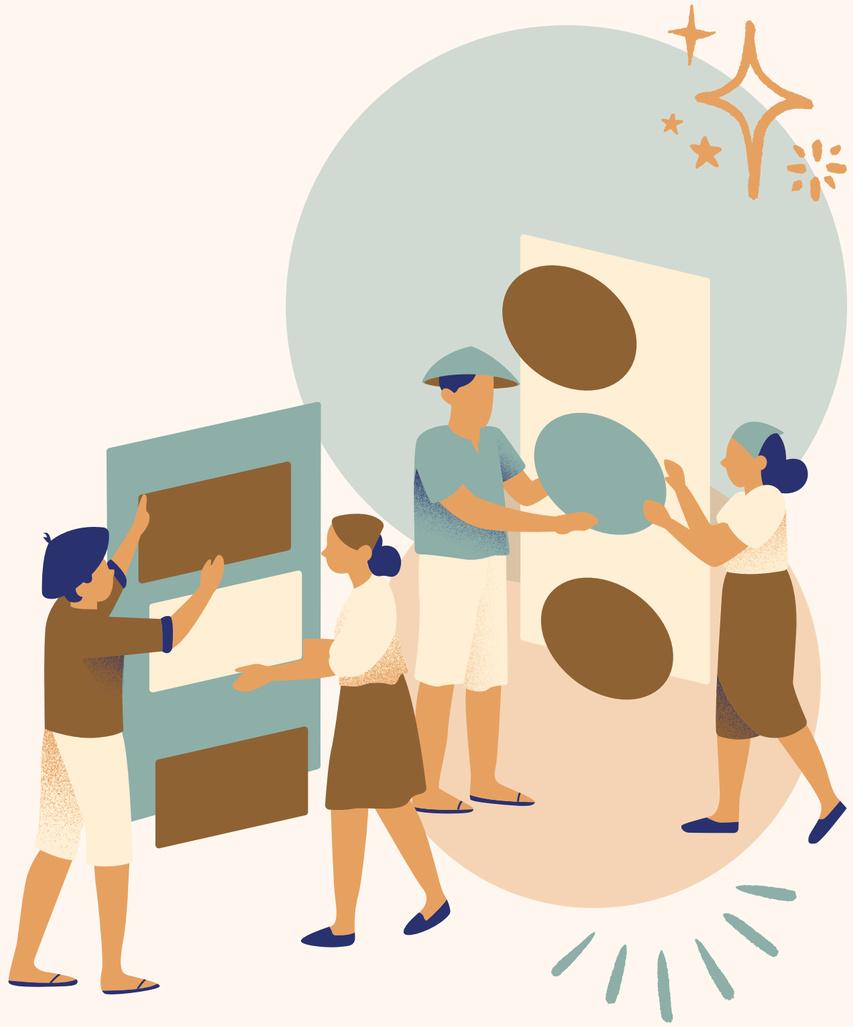


Project objective

Understanding job trends
(salary, job roles, location, work settings etc.)
in data science industry for job seekers

Problem to solve

- Automate the management job-related data.
- Store, manupulate & analyze job data efficiently
- Identify trends in job data more easily
- Validate the accuracy & consistancy of specific data



UNDERSTANDING *Data set*

Description

- CSV file(‘jobs_in_data.csv’) coming from https://www.kaggle.com/datasets/hummamqaasim/jobs-in-data & storing:

JOB-RELATED ATTRIBUTES

- **job_title:** Specifies the role (e.g., Data Scientist, Data Engineer).
- **job_category**: Groups job titles into broader categories like Machine Learning, Data Engineering, etc.
- **experience_level**: Categorizes experience (Entry-level, Mid-level, Senior, Executive).
- **employment_type**: Differentiates between Full-time, Part-time & Contract positions.

SALARY AND COMPENSATION

- **salary**: The annual gross salary in the local currency, offering raw salary data for regional comparisons.
- **salary_in_usd**: The salary converted to USD, enabling consistent global comparisons without the influence of currency fluctuations.
- **salary_currency**: (e.g., USD, EUR), important for currency conversion and understanding salary values globally.

GEOGRAPHIC INFORMATION

- **employee_residence**: The country where the employee resides, which helps analyze geographical salary differences.
- **company_location**: The country where the company is located, useful for understanding the impact of company location on salary structures.

WORK ENVIRONMENT

- **work_year:** The year the salary data was recorded, helping to track trends over time.
- **work_setting:** Defines whether the role is remote, hybrid, or in-person, providing insights into how different work settings impact salary levels.
- **company_size:** The size of the company, which can influence salary structures (e.g., larger companies may offer higher salaries).

sample data

```
2023,Data DevOps Engineer,Data Engineering,EUR,88000,95012,Germany,Mid-level,Full-time,Hybrid,Germany,L
2023,Data Architect,Data Architecture and Modeling,USD,186000,186000,United States,Senior,Full-time,In-person,United States,M
2023,Data Architect,Data Architecture and Modeling,USD,81800,81800,United States,Senior,Full-time,In-person,United States,M
```

PROGRAM CAPABILITIES

Features

CRUD method

- Read data
- Update data
- Create Data
- Delete Data

```
【TOP Menu】
1. Read all data
2. Read 50 row data
3. Creat a new record
4. Update an existing record
5. Delete a worker record
6. Basic analyzer
7. Salary analyzer
8. Exit
Enter your choice:
```

Basic Analyzer

- Null check
- Category check
- Frequency
- Sort
- Convert

```
【Basic Analysis Menu】
1. Find null
2. Check category
3. Search value
4. Check frequency
5. Convert salary currency from USD to HUF
6. Sort all values
7. Sort category
8. Sort frequency list
9. Go back to prior menu
Enter your choice:
```

Salary Analyzer

- Average
- Min/Max
- Min/Max by job
- Average by experience

```
【Salary Analyzer Menu】
1. Show AVERAGE salary
2. Show MAX salary
3. Show MIN salary
4. Show MAX salary job
5. Show MIN salary job
6. Average salary by experience
7. Go back to prior menu
Enter your choice:
```

EXAMPLE & RESULT- CRUD OPERATIONS

read data

- **Output: display the total number of records and all the records from the dataset**

```
【TOP Menu】
1. Read all data
2. Read 50 row data
3. Creat a new record
4. Update an existing record
5. Delete a worker record
6. Basic analyzer
7. Salary analyzer
8. Exit
Enter your choice: 1
Total records: 9355
Row 1:  {'work_year': '2023', 'job_title': 'Data DevOps Engineer', 'job_category': 'Data Engi
residence': 'Germany', 'experience_level': 'Mid-level', 'employment_type': 'Full-time', 'work

Row 2:  {'work_year': '2023', 'job_title': 'Data Architect', 'job_category': 'Data Architectu
'employee_residence': 'United States', 'experience_level': 'Senior', 'employment_type': 'Full
M'}

Row 3:  {'work_year': '2023', 'job_title': 'Data Architect', 'job_category': 'Data Architectu
mployee_residence': 'United States', 'experience_level': 'Senior', 'employment_type': 'Full-t
}

Row 4:  {'work_year': '2023', 'job_title': 'Data Scientist', 'job_category': 'Data Science an
oyee_residence': 'United States', 'experience_level': 'Senior', 'employment_type': 'Full-time

Row 5:  {'work_year': '2023', 'job_title': 'Data Scientist', 'job_category': 'Data Science an
ee_residence': 'United States', 'experience_level': 'Senior', 'employment_type': 'Full-time',
```

read first 50 rows

- **Output: dispay the first 50 records from the dataset**

```
Enter your choice: 2
Read first 50 rows
[{'work_year': '2023', 'job_title': 'Data DevOps Engineer', 'job_category': 'Data Engineering', 'salary_currency': 'EUR', 'salary': '88000', 'salary_in_usd': '95012', 'employee_residen
ce': 'Germany', 'experience_level': 'Mid-level', 'employment_type': 'Full-time', 'work_setting': 'Hybrid', 'company_location': 'Germany', 'company_size': 'L'}, {'work_year': '2023', 'j
ob_title': 'Data Architect', 'job_category': 'Data Architecture and Modeling', 'salary_currency': 'USD', 'salary': '186000', 'salary_in_usd': '186000', 'employee_residence': 'United St
ates', 'experience_level': 'Senior', 'employment_type': 'Full-time', 'work_setting': 'In-person', 'company_location': 'United States', 'company_size': 'M'}, {'work_year': '2023', 'job_
title': 'Data Architect', 'job_category': 'Data Architecture and Modeling', 'salary_currency': 'USD', 'salary': '81800', 'salary_in_usd': '81800', 'employee_residence': 'United States'
, 'experience_level': 'Senior', 'employment_type': 'Full-time', 'work_setting': 'In-person', 'company_location': 'United States', 'company_size': 'M'}, {'work_year': '2023', 'job_title
': 'Data Scientist', 'job_category': 'Data Science and Research', 'salary_currency': 'USD', 'salary': '212000', 'salary_in_usd': '212000', 'employee_residence': 'United States', 'exper
ience_level': 'Senior', 'employment_type': 'Full-time', 'work_setting': 'In-person', 'company_location': 'United States', 'company_size': 'M'}, {'work_year': '2023', 'job_title': 'Data
 Scientist', 'job_category': 'Data Science and Research', 'salary_currency': 'USD', 'salary': '93300', 'salary_in_usd': '93300', 'employee_residence': 'United States', 'experience_leve
l': 'Senior', 'employment_type': 'Full-time', 'work_setting': 'In-person', 'company_location': 'United States', 'company_size': 'M'}, {'work_year': '2023', 'job_title': 'Data Scientist
', 'job_category': 'Data Science and Research', 'salary_currency': 'USD', 'salary': '130000', 'salary_in_usd': '130000', 'employee_residence': 'United States', 'experience_level': 'Sen
ior', 'employment_type': 'Full-time', 'work_setting': 'Remote', 'company_location': 'United States', 'company_size': 'M'}, {'work_year': '2023', 'job_title': 'Data Scientist', 'job_cat
egory': 'Data Science and Research', 'salary_currency': 'USD', 'salary': '100000', 'salary_in_usd': '100000', 'employee_residence': 'United States', 'experience_level': 'Senior', 'empl
oyment_type': 'Full-time', 'work_setting': 'Remote', 'company_location': 'United States', 'company_size': 'M'}, {'work_year': '2023', 'job_title': 'Machine Learning Researcher', 'job_c
ategory': 'Machine Learning and AI', 'salary_currency': 'USD', 'salary': '224400', 'salary_in_usd': '224400', 'employee_residence': 'United States', 'experience_level': 'Mid-level', 'e
mployment_type': 'Full-time', 'work_setting': 'In-person', 'company_location': 'United States', 'company_size': 'M'}, {'work_year': '2023', 'job_title': 'Machine Learning Researcher',
 'job_category': 'Machine Learning and AI', 'salary_currency': 'USD', 'salary': '138700', 'salary_in_usd': '138700', 'employee_residence': 'United States', 'experience_level': 'Mid-leve
l', 'employment_type': 'Full-time', 'work_setting': 'In-person', 'company_location': 'United States', 'company_size': 'M'}, {'work_year': '2023', 'job_title': 'Data Engineer', 'job_cat
egory': 'Data Engineering', 'salary_currency': 'USD', 'salary': '210000', 'salary_in_usd': '210000', 'employee_residence': 'United States', 'experience_level': 'Executive', 'employment
_type': 'Full-time', 'work_setting': 'Remote', 'company_location': 'United States', 'company_size': 'M'}, {'work_year': '2023', 'job_title': 'Data Engineer', 'job_category': 'Data Engi
neering', 'salary_currency': 'USD', 'salary': '168000', 'salary_in_usd': '168000', 'employee_residence': 'United States', 'experience_level': 'Executive', 'employment_type': 'Full-time
', 'work_setting': 'Remote', 'company_location': 'United States', 'company_size': 'M'}, {'work_year': '2023', 'job_title': 'Machine Learning Engineer', 'job_category': 'Machine Learnin
g and AI', 'salary_currency': 'USD', 'salary': '224400', 'salary_in_usd': '224400', 'employee_residence': 'United States', 'experience_level': 'Senior', 'employment_type': 'Full-time',
 'work_setting': 'In-person', 'company_location': 'United States', 'company_size': 'M'}, {'work_year': '2023', 'job_title': 'Machine Learning Engineer', 'job_category': 'Machine Learni
ng and AI', 'salary_currency': 'USD', 'salary': '138700', 'salary_in_usd': '138700', 'employee_residence': 'United States', 'experience_level': 'Senior', 'employment_type': 'Full-time'
, 'work_setting': 'In-person', 'company_location': 'United States', 'company_size': 'M'}, {'work_year': '2023', 'job_title': 'Data Scientist', 'job_category': 'Data Science and Researc
h', 'salary_currency': 'GBP', 'salary': '35000', 'salary_in_usd': '43064', 'employee_residence': 'United Kingdom', 'experience_level': 'Mid-level', 'employment_type': 'Full-time', 'wor
k_setting': 'In-person', 'company_location': 'United Kingdom', 'company_size': 'M'}, {'work_year': '2023', 'job_title': 'Data Scientist', 'job_category': 'Data Science and Research', '
salary_currency': 'GBP', 'salary': '30000', 'salary_in_usd': '36912', 'employee_residence': 'United Kingdom', 'experience_level': 'Mid-level', 'employment_type': 'Full-time', 'work_set
ting': 'In-person', 'company_location': 'United Kingdom', 'company_size': 'M'}, {'work_year': '2023', 'job_title': 'Data Analyst', 'job_category': 'Data Analysis', 'salary_currency': '
USD', 'salary': '95000', 'salary_in_usd': '95000', 'employee_residence': 'United States', 'experience_level': 'Entry-level', 'employment_type': 'Full-time', 'work_setting': 'In-person'
, 'company_location': 'United States', 'company_size': 'M'}, {'work_year': '2023', 'job_title': 'Data Analyst', 'job_category': 'Data Analysis', 'salary_currency': 'USD', 'salary': '75
000', 'salary_in_usd': '75000', 'employee_residence': 'United States', 'experience_level': 'Entry-level', 'employment_type': 'Full-time', 'work_setting': 'In-person', 'company_location
': 'United States', 'company_size': 'M'}, {'work_year': '2023', 'job_title': 'Data Scientist', 'job_category': 'Data Science and Research', 'salary_currency': 'USD', 'salary': '300000'
, 'salary_in_usd': '300000', 'employee_residence': 'United States', 'experience_level': 'Senior', 'employment_type': 'Full-time', 'work_setting': 'In-person', 'company_location': 'Unit
ed States', 'company_size': 'M'}, {'work_year': '2023', 'job_title': 'Data Scientist', 'job_category': 'Data Science and Research', 'salary_currency': 'USD', 'salary': '234000', 'salar
```

EXAMPLE & RESULT- CRUD OPRRATIONS

create data

- Input: column names and values for the record you would like to create
- Outputs: display a message confirming the creation

```
【TOP Menu】
1. Read all data
2. Read 50 row data
3. Creat a new record
4. Update an existing record
5. Delete a worker record
6. Basic analyzer
7. Salary analyzer
8. Exit
Enter your choice: 3
plase enter a new record
for example
{'work_year': '2024', 'job_title': 'Data Scientist', 'job_category': 'Data Science', 'salary_currency': 'USD','salary':'120000', 'salary_in_usd': '120000',
'employee_residence': 'United States', 'experience_level': 'Senior', 'employment_type': 'Full-time', 'work_setting': 'Remote', 'company_location': 'United S
tates', 'company_size': 'M'}
enter here: {'work_year': '2024', 'job_title': 'Data Scientist', 'job_category': 'Data Science', 'salary_currency': 'USD','salary':'120000', 'salary_in_usd
: '120000', 'employee_residence': 'United States', 'experience_level': 'Senior', 'employment_type': 'Full-time', 'work_setting': 'Remote', 'company_locatio
': 'United States', 'company_size': 'M'}
successful saved
```

update data

- Input: column names and values for the record you would like to update
- Outputs: display a message confirming the update / error

```
Enter your choice: 4
which column and value do you want to chage? ex) {'job_title': 'Data Science',
 'work_year': '2024'} -> {'experience_level': 'Lead'}
target data: {'work_year':'2024'}
update data: {'work_year':'2023'}
Record updated with {'work_year': '2023'}
```

```
Enter your choice: 4
which column and value do you want to chage? ex) {'job_title': 'Data Science', 'work_year': '2024'} -> {'experience_level': 'Lea
target data: {'job_title': 'Data Science', 'work_year': '2024'}
update data: {'experience_level': 'Lead'}
No record found with the specified attributes.
```

delete data

- Input: column names and values for the record you would like to delete
- Output: display a message confirming the deletion / error

```
Enter your choice: 5
Which data do you want to delete? ex){'job_title': 'Data Scientist', 'work_year': '2024'}
target data: {'job_title': 'Data Scientist', 'work_year': '2024'}
Record deleted.
```

EXAMPLE & RESULT- BASIC ANALYZER

check null

- Input: column name you want to check (company_location)
- Output:
  - if exists -> tell which row it is in
  - if not -> tell there is no null value in the column

check category

- Input: column name (experience_level)
- Output: the unique value in the column as a set
  - (Entry-level, Senior, Mid-level, Executive)

search value

- Input: column name & a specific word for serching (work_year, 2020)
- Output:
  - if exists -> tell which row it is in as a list
  - if not -> tell you there is no such a data

```
【Basic Analysis Menu】
1. Find null
2. Check category
3. Search value
4. Check frequency
5. Convert salary currency from USD to HUF
6. Sort all values
7. Sort category
8. Sort frequency list
9. Go back to prior menu
Enter your choice: 1
Please enter column name: company_location
there is no null value
```

```
Enter your choice: 2
Please enter column name: experience_level
{'Entry-level', 'Senior', 'Mid-level', 'Exec
```

```
Enter your choice: 3
Please enter column name: work_year
Please enter word you want to serch: 2020
the value (2020) is in the [2086, 2289, 4115, 4592, 46
80, 9185, 9190, 9193, 9202, 9203, 9205, 9208, 9211, 92
92, 9294, 9297, 9298, 9304, 9306, 9311, 9313, 9315, 93
```

```
Enter your choice: 3
Please enter column name: work_year
Please enter word you want to serch: 2025
there isn't such a data (2025)
```

EXAMPLE & RESULT- BASIC ANALYZER

frequency

- Input: column name (company_location)
- Output: frequency list

```
Enter your choice: 4
Please enter column name: work_year
{'2023': 7453, '2022': 1634, '2020': 71, '2021': 197}
```

change currency

- Input: currency rate from $USD to some currency you want ($1(USD)=390(HUF))
- Output: all salary data converted to HUF

```
Enter your choice: 5
Please enter currency rate: 390
[37054680.0, 72540000.0, 31902000.0, 82680000.0, 36387000.0, 50700
50000.0, 29250000.0, 117000000.0, 91260000.0, 54600000.0, 46800000
8000.0, 65325000.0, 41535000.0, 72501000.0, 50427000.0, 72150000.0
```

sort (by all values / category / frequency)

- Input:
  - by all values / category -> column name & asc /desc order
  - by frequency -> column name, asc /desc order & key / value
- Output: sorted list

```
Enter your choice: 7
Please enter column name: work_year
Please enter asc or desc: asc
['2020', '2021', '2022', '2023']
```

```
Enter your choice: 8
Please enter column name: work_year
Please enter asc or desc: asc
Please enter key or value: value
{'2020': 71, '2021': 197, '2022': 1634, '2023': 7453}
```

EXAMPLE & RESULT- SALARY ANALYZER

- **Input: numbers from Salary Analyzer Menu**
- **Outputs:**

```
【Salary Analyzer Menu】
1. Show AVERAGE salary
2. Show MAX salary
3. Show MIN salary
4. Show MAX salary job
5. Show MIN salary job
6. Average salary by experience
7. Go back to prior menu
Enter your choice: 2
The maximum salary in USD is: 450000.0
```

```
Enter your choice: 1
The average salary in USD is: 150299.5
Total salary of 9355 employees is 1406051781.0
```

```
Enter your choice: 3
The minimum salary in USD is: 15000.0
```

```
Enter your choice: 4
The job with the highest salary is: Research Scientist with a salary of 450000.0 USD
```

```
Enter your choice: 5
The job with the minimum salary is: Business Intelligence Developer with a salary of 15000.0 USD
```

```
Enter your choice: 6
Employee segmentation by experience level with their average salary:
Mid-level: 117523.92
Senior: 162356.13
Executive: 189462.91
Entry-level: 88534.78
```

CONCLUSION AND FUTURE WORK

Future Enhancements:

Adding a user-friendly interface.

Expanding capabilities to include statistical analysis

Conclusion & future work

- CRUD
- TEXT FILE METHODS
- BASIC ALGORITHMS (BASIC ANALYZER)
- STATISTICAL BUSINESS CALCULATIONS (SALARY ANALYZER)
- Main program
- Testing

THANK *You*