

PROJECT: ANALYSIS OF EMPLOYEE SALARY DATA DATASET: Data.csv	PYTHON PROGRAMMING I	LEVEL: PI1 A & B & C
---	-----------------------------	-------------------------------------

PART I — File Handling

(Manual processing, no Pandas allowed.)

1. Preview the dataset
 2. Search for “chief” in the CSV
 3. Extract manually 2 columns
-

PART II — Data Cleaning (Pandas)

4. Clean the dataset
 - Replace missing numeric with 0
 - Remove duplicates
 - ...
-

PART III — Subsetting & Filtering

5. Create subsets based on relevant and well-defined criteria
 - a) High earners
 - b) Employees from year 2013
 - c) JobTitle contains “POLICE”
 - ...
-

PART IV — New Columns & Summary Statistics

6. Create computed columns
 - **Is_Manager = True/False** if JobTitle contains "MANAGER" or "CHIEF"
 - ...
7. Create a summary statistics
 - Average BasePay
 - Top 5 most common titles

- Total number of employees
 - ...
-

PART V — Grouping & Aggregation

8. Perform an analysis using group-based aggregation

- Average TotalPay per Year
 - ...
-

PART VI — Joining

9. Merge with `agency_codes.csv`

- Create a mapping file
 - Merge main dataset + mapping on “Agency”
 - Save as `merged_data.csv`
-

PART VII — Interactive Investigation Script

10. `investigation.py`

- Ask user for job title keyword
- Filter rows
- Show:
 - number of matches
 - average BasePay
 - highest TotalPay
- Save results → `custom_search.csv`
- Handle errors cleanly