

Mid-term Exam

Full name 1 (Student ID Number 1)

Full name 2 (Student ID Number 2) (optional)

Note:

- You may answer and present this problem using English or Bahasa Indonesia.
- This exam is expected to take approximately 10-15 hours in total to complete.

Problem Background

Recent discussions have raised questions about the accuracy and calculation methods of Indonesia's national inflation statistics. Comment by Mr. Tito Karnavian during the "Anugrah Hari Statistik Nasional 2024" event ([YouTube link](#), from 1:52:16 to 1:55:00) and suspicious from economists reported in a [Bloomberg article](#) (you can open the link with the help of [Remove Paywall](#)) suggest potential discrepancies in the reported figures.

Tasks

1. Data Collection

- **Provincial Data:** Collect monthly inflation data for each province in Indonesia for all available years and months the data should be formatted as follows:

province-name.csv

```
date,inflation_percent
...
2023-01,0.43
2023-02,0.11
2023-03,0.59
2023-04,0.42
...
```

- An example dataset for East Kalimantan can be found [here](#).
 - At the end you will have 38 **.csv** files.
- **National Data:** Obtain the national inflation data from the [Inflation \(in percentage\), 2024 - Indonesia](#)

2. Data Processing with PySpark

- **Load Data:** Use PySpark to read the provincial and national inflation data.
- **Data Cleaning:** Ensure the data is clean, properly formatted, and handle any missing values.
- **Aggregation:** Calculate the mean inflation rate for each month across all provinces.
- **Comparison:** Compare the aggregated provincial mean inflation rates with the national inflation rates.

- **Calculation of Differences:** Compute the absolute differences between the provincial mean inflation rates and the national inflation rates for each month.

3. Analysis

- **Investigate Discrepancies:** Analyze whether the national inflation rate is simply the mean of the provincial rates.
- **Identify Patterns:** Look for any patterns or anomalies in the data over time.
- **Provide Explanation:** Offer possible explanations for any discrepancies found, referencing economic principles or data reporting practices.

4. Reporting

- **Markdown Report** (`report.md`): Document your work, including:
 - Introduction and motivation.
 - Methodology for data collection and processing.
 - Results with supporting graphs and tables.
 - Discussion and interpretation of findings. You can relate to the possibilities of creating a new business idea from this problem for example being as a third-party who provides data calibration and data collecting more reliable. Elaborate with that ideas, impact and consequences.
 - Conclusion summarizing your insights.
- **Ensure Clarity:** Write clearly and concisely, avoiding plagiarism and properly citing any sources used.

3. Presentation

- **Slide** (`slides.pdf` or `slides.ppts`): Create a visual presentation summarizing your project.
- **Video Presentation:** Record a video (maximum 15 minutes) presenting your work. The presentation should:
 - Introduce the problem and its significance.
 - Explain your methodology and analysis.
 - Highlight key findings and conclusions.
 - Be engaging and avoid reading directly from notes.

Submission Guidelines

You may work individually or with one partner. Organize your submission with the following directory structure:

```
StudentID1-StudentID2-midterm-exam
├── data
│   ├── indonesia.csv
│   ├── province-name01.csv
│   ├── province-name02.csv
│   ├── province-name03.csv
│   └── ...
├── program
│   └── indonesia_inflation_analysis.ipynb
├── report
│   └── report.md
```

```

├── image1.png (optional)
├── slides
│   ├── slides.pdf (or .pptx)
└── video
    └── presentation.mp4

```

- **data**: Contains all the datasets used. You must change the name `province-name01.csv` according to the name of the province name where the data is collected.
- **program**: Your PySpark Jupyter Notebook.
- **report**: Your written report and any additional images.
- **slides**: Your presentation slides.
- **video**: Your recorded presentation.

Evaluation Rubric

Points	Criteria
10	[S] Present confidently with natural speech, appropriate facial expressions, and eye contact, demonstrating understanding without reading from notes
10	[S,R] Show a clear understanding of the problem and outlines an effective approach to solve it
10	[S,R,P] Ensures that the solution directly addresses the problem and is consistently presented across all components
10	[P] Coding structure is well-organized that follows the code presented during the class or the documentation of PySpark
10	[S] Uses slides that are simple, visually appealing, and contain concise text to support key points
10	[R] Provides correct answers in the report, written in the student's own words, demonstrating original thought and proper citation of sources
10	[S,R,P] Follow all provided instructions and submission guidelines accurately across all components.
30	[S,R,P] Submit work that is original and not copied from others, showing honesty and integrity in all components.

Abbreviation: [S]: Slide/Presentation, [R]: Report, and [P]: Program