

# Code Book

## Background

### Background\Education Level

- Background\Education Level\Master: 5
- Background\Education Level\PhD: 5

### Background\Graduation Department

- Background\Graduation Department\Computer Science: 6
- Background\Graduation Department\Statistics: 2
- Background\Graduation Department\Finance: 1
- Background\Graduation Department\Accounting: 1

### Background\Current Role

- Background\Current Role\Students: 4
- Background\Current Role\Data Engineer: 2
- Background\Current Role\Data Scientists: 1
- Background\Current Role\Software Engineer: 2
- Background\Current Role\Quantitative Analyst: 1

### Background\Years of Programming Experience

- Background\Years of Programming Experience\2-5 years: 2
- Background\Years of Programming Experience\5+ years: 8

### Background\Years of Data Science Experience

- Background\Years of Data Science Experience< 2 years: 1

- Background\Years of Data Science Experience\2-5 years: 4
- Background\Years of Data Science Experience\5+ years: 5

## **Background\Work Place**

- Background\Work Place\School: 4
- Background\Work Place\Tech Company: 3
- Background\Work Place\Insurance Company: 1
- Background\Work Place\Financial Company: 2

## **Background\Locations**

- Background\Locations\R1 University: 4
- Background\Locations\Boston: 1
- Background\Locations\New York: 1
- Background\Locations\Chicago: 1
- Background\Locations\Asia: 1

## **Root Causes**

### **Root Causes\Why Data Preparaion is difficult and error-prone**

- Root Causes\Why Data Preparaion is error-prone\Data is usually dirty (missing values, incomplete data, inconsisten data type): 9
- Root Causes\Why Data Preparaion is error-prone\The data's format is unclear: 8
- Root Causes\Why Data Preparaion is error-prone\Lack of domain knowledge: 5
- Root Causes\Why Data Preparaion is error-prone\Integrating different data sources is error-prone: 4
- Root Causes\Why Data Preparaion is error-prone\Data size is too large: 4

- Root Causes\Why Data Preparation is error-prone\Automating the data cleaning process is challenging: 3
- Root Causes\Why Data Preparation is error-prone\Transforming unstructured data into a meaningful format is difficult: 2
- Root Causes\Why Data Preparation is error-prone\Need to remember column names: 1

## **Root Causes\Why Data Exploration is difficult and error-prone**

- Root Causes\Why Data Exploration is error-prone\Unfamiliar with dataset: 6
- Root Causes\Why Data Exploration is error-prone\There are too many columns to visualize, handle, or choose from: 5
- Root Causes\Why Data Exploration is error-prone\It's difficult to determine which features are important: 4
- Root Causes\Why Data Exploration is error-prone\Wrong assumption about the data: 4
- Root Causes\Why Data Exploration is error-prone\Inconsistent data types: 3
- Root Causes\Why Data Exploration is error-prone\Lack of domain knowledge: 3
- Root Causes\Why Data Exploration is error-prone\Unfamiliar with third-party libraries and APIs: 1

## **Root Causes\Why Training model is difficult and error-prone**

- Root Causes\Why Training model is difficult and error-prone\Fine-tuning is time-consuming: 2

## **Root Causes\Debugging Challenges in Jupyter Notebooks**

- Root Causes\Debugging Challenges in Jupyter Notebooks\Need to memorize cell dependency: 6
- Root Causes\Debugging Challenges in Jupyter Notebooks\Large notebooks can create significant dependency issues: 6
- Root Causes\Debugging Challenges in Jupyter Notebooks\Complex variable dependency issues: 4

- Root Causes\Debugging Challenges in Jupyter Notebooks\Rerunning from scratch can be time-consuming if dependency issues occur: 3

# Debugging Practice

## Debugging Practice

- Debugging Practice\Use 'print': 9
- Debugging Practice\Switch between languages and tools: 8
- Debugging Practice\Search online: 6
- Debugging Practice\Organize code into functions and test: 5
- Debugging Practice\Execute cell individually to narrow down error: 3
- Debugging Practice\Rerun from the first cell: 3
- Debugging Practice\Read Error logs and trace back to the bug: 3
- Debugging Practice\Sample a smaller dataset for testing: 1
- Debugging Practice\Write notes in the markdown cells: 1

# Improvement and Opportunity

## Improvement and Opportunity

- Improvement and Opportunity\Integrate AI into notebooks: 6
- Improvement and Opportunity\Tools to manage/visualize cell dependencies: 5
- Improvement and Opportunity\Need version control tools for the 'state' rather than the code.: 4
- Improvement and Opportunity\Integrate mature third-party tools into notebooks: 4
- Improvement and Opportunity\Display the current variable value: 3
- Improvement and Opportunity\On-hover features display API usage and variable types: 3
- Improvement and Opportunity\Support for breakpoints in notebooks: 2

- Improvement and Opportunity\Automated typo correction feature: 1