Code List

Background	
	Background\Education Level
	Background\Graduation Department
	Background\Current Role
	Background\Years of Programming Experience
	Background\Years of Data Science Experience
Root Causes	
	Root Causes\Why Data Preparaion is difficult and error-pro
	Root Causes\Why Data Exploration is difficult and error-pro
	Root Causes\Why Training model is difficult and error-pron
	Root Causes\Cell Dependency

Debugging Practice

Debugging Practice

Improvement and Opportunity

Improvement and Opportunity

Background\Education Level\Master
Background\Education Level\PhD

Background\Graduation Department\Computer Science
Background\Graduation Department\Statistics
Background\Graduation Department\Finance
Background\Graduation Department\Accounting

Background\Current Role\Students
Background\Current Role\Data Engineer
Background\Current Role\Data Scientists
Background\Current Role\Software Engineer
Background\Current Role\Quantitative Analyst

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

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

one
Root Causes\Why Data Preparaion is error-prone\Data is usually dirty (missing values, incomplete data,
Root Causes\Why Data Preparaion is error-prone\The data's format is unclear
Root Causes\Why Data Preparaion is error-prone\Lack of domain knowledge
Root Causes\Why Data Preparaion is error-prone\Integrating different data sources is error-prone
Root Causes\Why Data Preparaion is error-prone\Data size is too large
Root Causes\Why Data Preparaion is error-prone\Automating the data cleaning process is challenging
Root Causes\Why Data Preparaion is error-prone\Transforming unstructured data into a meaningful form
one
Root Causes\Why Data Exploration is error-prone\Unfamiliar with dataset
Root Causes\Why Data Exploration is error-prone\There are too many columns to visualize, handle, or c
Root Causes\Why Data Exploration is error-prone\It's difficult to determine which features are important
Root Causes\Why Data Exploration is error-prone\Wrong assumption about the data
Root Causes\Why Data Exploration is error-prone\Inconsistent data types
Root Causes\Why Data Exploration is error-prone\Lack of domain knowledge
ne
Root Causes\Why Training model is difficult and error-prone\Fine-tuning is time-consuming

Root Causes\Cell Dependency\Large notebooks can create significant dependency issues

Root Causes\Cell Dependency\Need to memorize cell dependency

Root Causes\Cell Dependency\Complex variable dependency issues

Root Causes\Cell Dependency\Rerunning from scratch can be time-consuming if dependency issues oc

Debugging Practice\Use 'print'

Debugging Practice\Search online

Debugging Practice\Organize code into functions and test

Debugging Practice\Execute cell individually to narrow down error

Debugging Practice\Rerun from the first cell

Debugging Practice\Read Error logs and trace back to the bug

Improvement and Opportunity\Tools to manage/visualize cell dependencies

Improvement and Opportunity\Need version control tools for the 'state' rather than the code.

Improvement and Opportunity\Display the current variable value

Improvement and Opportunity\On-hover features display API usage and variable types

Improvement and Opportunity\Support for breakpoints in notebooks

Times
5
5
6
2
1
1
4
2
1
2
1
2
8
1
4
5
9
8
5
4
4
3
2
6
5
4
4
3
3
2
6

6
4
3

9
6
5
3
3
3

5
4
3
3
2