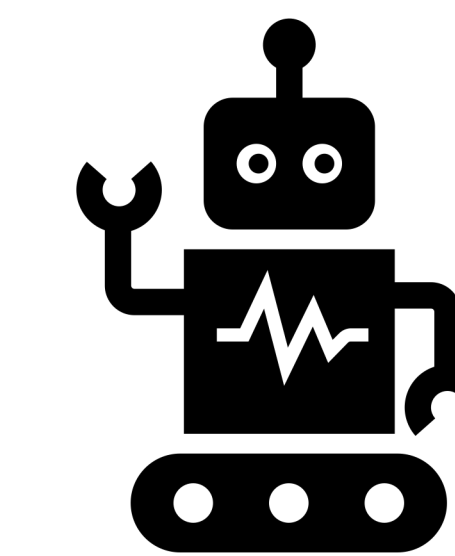
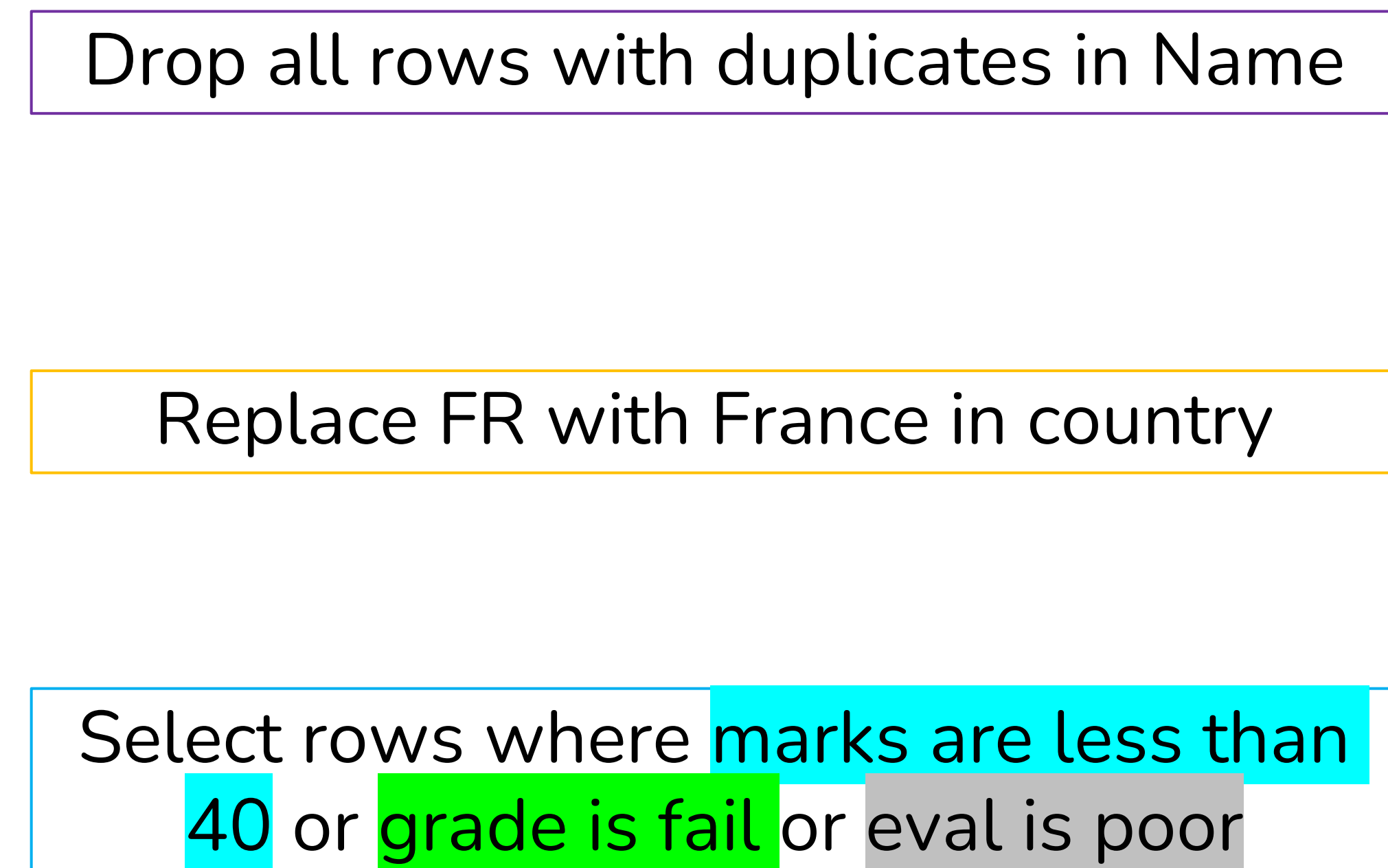
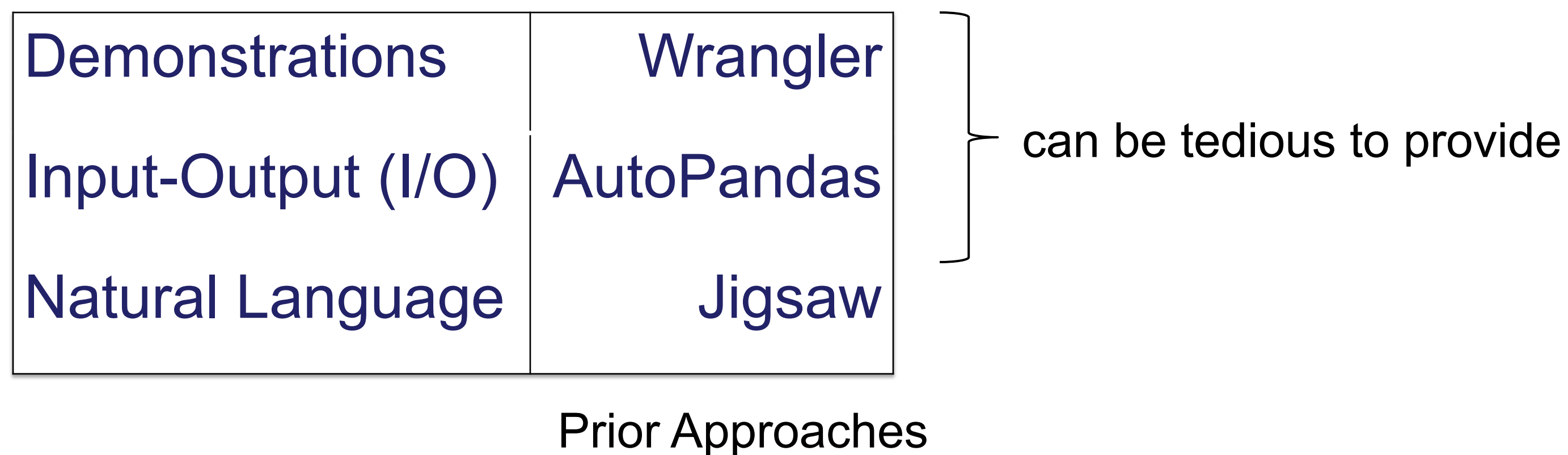
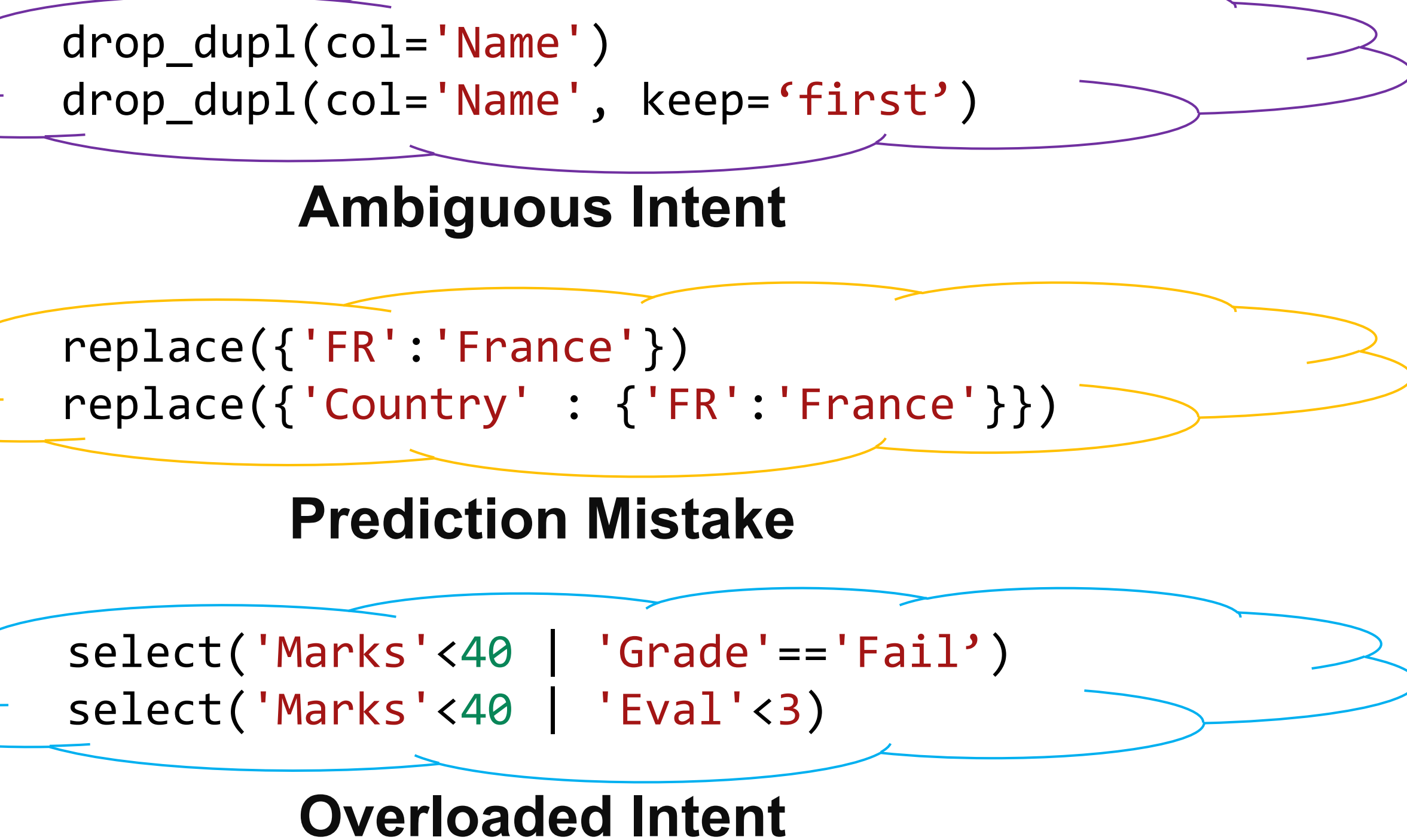


Motivation and Prior Work

Data Wrangling requires users to perform ad-hoc operations in complex languages like Pandas with over 700 API functions, each often requiring 3-5 arguments, making it difficult for the users to utilize these APIs effectively



Issues with natural language to code systems

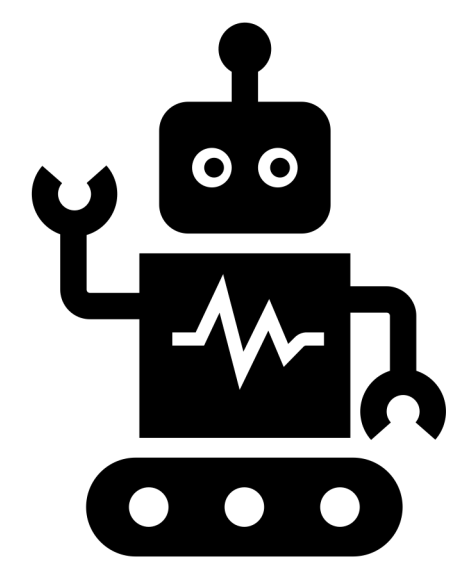


Multi-Modal Interactivity for Grounding and Debugging

Key Idea: Users begin specifying intent using natural language and can further interact with the system to refine specification or fix mistakes

Natural Language Interaction

User Query: Drop all rows with duplicates in Name



Code2NL

- lexically similar summaries to original query

Filter all the rows having duplicates in column Name

bad summary

Drop all rows with duplicates in column Name including the first duplicate

good summary

Recommend Alternatives

drop_dupl(col='Name')
drop_dupl(col='Name', keep='first')

Alternatively consider:
Drop all rows with duplicates in Name while keeping the first duplicate

Disambiguate Alternatives

Clarification:
Do you want to drop all duplicate rows or keep the first duplicate?

semantic difference

Natural Language Interaction

Multi-Modal Interaction

Compositional Interaction

Interaction Modules

Multi-Modal Interaction

replace({'FR': 'France'})

Country	Capital	Origin
Germany	Berlin	USA
India	Delhi	FR-France
FR-France	Paris	UK

Do you want to replace this?

select('Score' > 15) select('Score' >= 15)

Name	Score
u0	42
u1	15
u2	10
u3	20

user selects another row

Name	Score
u0	42
u1	15
u2	10
u3	20

Repair by Demonstrations

Compositional Interaction

Decompose into subproblems:

- Select rows where marks are less than 40
- Select rows where grade is fail
- Select rows where eval is poor
- Select all the rows from above