

Lesson 5:
Data Analysis Process - Case Study 2

SEARCH

RESOURCES

CONCEPTS

5. Assessing Data

6. Cleaning Column Labels

7. Filter, Drop Nulls, Dedupe

8. Inspecting Data Types

9. Fixing Data Types Pt 1

10. Fixing Data Types Pt 2

11. Fixing Data Types Pt 3

12. Exploring with Visuals

13. Conclusions & Visuals

14. Types of Merges

15. Merging Datasets

16. Results with Merged Dataset

17. Conclusion

Lesson 5:
Data Analysis Process - Case Study 2

SEARCH

RESOURCES

CONCEPTS

5. Assessing Data

6. Cleaning Column Labels

7. Filter, Drop Nulls, Dedupe

8. Inspecting Data Types

9. Fixing Data Types Pt 1

10. Fixing Data Types Pt 2

11. Fixing Data Types Pt 3

12. Exploring with Visuals

13. Conclusions & Visuals

14. Types of Merges

15. Merging Datasets

16. Results with Merged Dataset

17. Conclusion

Lesson 5:
Data Analysis Process - Case Study 2

SEARCH

RESOURCES

CONCEPTS

5. Assessing Data

6. Cleaning Column Labels

7. Filter, Drop Nulls, Dedupe

8. Inspecting Data Types

9. Fixing Data Types Pt 1

10. Fixing Data Types Pt 2

11. Fixing Data Types Pt 3

12. Exploring with Visuals

13. Conclusions & Visuals

14. Types of Merges

15. Merging Datasets

16. Results with Merged Dataset

17. Conclusion

Lesson 5:
Data Analysis Process - Case Study 2

SEARCH

RESOURCES

CONCEPTS

5. Assessing Data

6. Cleaning Column Labels

7. Filter, Drop Nulls, Dedupe

8. Inspecting Data Types

9. Fixing Data Types Pt 1

10. Fixing Data Types Pt 2

11. Fixing Data Types Pt 3

12. Exploring with Visuals

13. Conclusions & Visuals

14. Types of Merges

15. Merging Datasets

16. Results with Merged Dataset

17. Conclusion

Lesson 5:
Data Analysis Process - Case Study 2

SEARCH

RESOURCES

CONCEPTS

5. Assessing Data

6. Cleaning Column Labels

7. Filter, Drop Nulls, Dedupe

8. Inspecting Data Types

9. Fixing Data Types Pt 1

10. Fixing Data Types Pt 2

11. Fixing Data Types Pt 3

12. Exploring with Visuals

13. Conclusions & Visuals

14. Types of Merges

15. Merging Datasets

16. Results with Merged Dataset

17. Conclusion

Lesson 5:
Data Analysis Process - Case Study 2

SEARCH

RESOURCES

CONCEPTS

5. Assessing Data

6. Cleaning Column Labels

7. Filter, Drop Nulls, Dedupe

8. Inspecting Data Types

9. Fixing Data Types Pt 1

10. Fixing Data Types Pt 2

11. Fixing Data Types Pt 3

12. Exploring with Visuals

13. Conclusions & Visuals

14. Types of Merges

15. Merging Datasets

16. Results with Merged Dataset

17. Conclusion

Lesson 5:
Data Analysis Process - Case Study 2

SEARCH

RESOURCES

CONCEPTS

5. Assessing Data

6. Cleaning Column Labels

7. Filter, Drop Nulls, Dedupe

8. Inspecting Data Types

9. Fixing Data Types Pt 1

10. Fixing Data Types Pt 2

11. Fixing Data Types Pt 3

12. Exploring with Visuals

13. Conclusions & Visuals

14. Types of Merges

15. Merging Datasets

16. Results with Merged Dataset

17. Conclusion

Assessing Data

SEND FEEDBACK

Assessing Data

The files `all_alpha_08.csv` and `all_alpha_18.csv` discussed in the previous pages have been provided in the workspace for you here to access. Use pandas to explore these datasets in the Jupyter Notebook below to answer the quiz questions below the notebook about these characteristics of the data:

- number of samples in each dataset
- number of columns in each dataset
- duplicate rows in each dataset
- datatypes of columns
- features with missing values
- number of non-null unique values for features in each dataset
- what those unique values are and counts for each

jupyter assessing (autosaved)

File Edit View Insert Cell Kernel Help Not Trusted Python 3

Assessing

Use the space below to explore `all_alpha_08.csv` and `all_alpha_18.csv` to answer the quiz questions below.

In []:

Assessing Data

SEND FEEDBACK

QUESTION 1 OF 6

Find the correct count for each of the following in the 2008 dataset

261611413889

FEATURE	COUNT
samples	2404
columns	18
duplicate rows	25
rows with missing data	199

SUBMIT

Assessing Data

SEND FEEDBACK

QUESTION 2 OF 6

Find the correct count for each of the following in the 2018 dataset

32152404

FEATURE	COUNT
samples	1611
columns	18
duplicate rows	0
rows with missing data	2

SUBMIT

QUESTION 3 OF 6

Match the datatype for each feature (some of these may not be ideal)

intboolboolfloat

Cyl (2018)	float
Cyl (2008)	string
City MPG (both)	string
Greenhouse Gas Score (2008)	string
Greenhouse Gas Score (2018)	int

SUBMIT

Assessing Data

SEND FEEDBACK

QUESTION 4 OF 6

Match the number of non-null unique values for each of the following features

181531434222

FEATURE	UNIQUE VALUES
SmartWay (2008)	

Assessing Data

SEND FEEDBACK

Sales Area (2008)	
Cert Region (2018)	
Trans (2008)	

SUBMIT

QUESTION 5 OF 6

How do the `Cyl` columns in the 2008 and 2018 datasets differ?

☒ Datatypes

☒ Format

☒ Number of unique values

☐ Meaning

Assessing Data

SEND FEEDBACK

QUESTION 6 OF 6

Where are each of these fuel types present?

2008NeitherBoth

FUEL TYPE	DATASET
Gasoline	Both
Ethanol/Gas	Both
Gasoline/Electricity	2018
CNG	2008
Electricity	2018

SUBMIT

NEXT