Economic Factors

Mariam Ahmad
Yuri Groza
Kasey Wilson

Data Sources

We selected two CSV files from Kaggle

US Gasoline and Diesel Retail Prices 1995-2021

US Unemployment Rate 1948-2019

Extract

Imported CSV files into two separate DataFrames

Unemployment

	Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
0	1948	3.4	3.8	4.0	3.9	3.5	3.6	3.6	3.9	3.8	3.7	3.8	4.0
1	1949	4.3	4.7	5.0	5.3	6.1	6.2	6.7	6.8	6.6	7.9	6.4	6.6
2	1950	6.5	6.4	6.3	5.8	5.5	5.4	5.0	4.5	4.4	4.2	4.2	4.3
3	1951	3.7	3.4	3.4	3.1	3.0	3.2	3.1	3.1	3.3	3.5	3.5	3.1
4	1952	3.2	3.1	2.9	2.9	3.0	3.0	3.2	3.4	3.1	3.0	2.8	2.7

Fuel

	Date	A1	A2	A3	R1	R2	R3	M1	M2	M3	P1	P2	P3	D1
0	01/02/1995	1.127	1.104	1.231	1.079	1.063	1.167	1.170	1.159	1.298	1.272	1.250	1.386	1.104
1	01/09/1995	1.134	1.111	1.232	1.086	1.070	1.169	1.177	1.164	1.300	1.279	1.256	1.387	1.102
2	01/16/1995	1.126	1.102	1.231	1.078	1.062	1.169	1.168	1.155	1.299	1.271	1.249	1.385	1.100
3	01/23/1995	1.132	1.110	1.226	1.083	1.068	1.165	1.177	1.165	1.296	1.277	1.256	1.378	1.095
4	01/30/1995	1.131	1.109	1.221	1.083	1.068	1.162	1.176	1.163	1.291	1.275	1.255	1.370	1.090

Transform

Unemployment DataFrame

- We converted columns to rows to mimic formatting of Fuel DataFrame
- Combined Month and Year columns to Month/Year to create Key
- Converted Month Formatting to Numeric

	year	month	rate	monthn	monthyear
0	1948	Jan	3.4	1	11948
1	1949	Jan	4.3	1	11949
2	1950	Jan	6.5	1	11950
3	1951	Jan	3.7	1	11951
4	1952	Jan	3.2	1	11952
...	...	...	...	...	...
859	2015	Dec	5.0	12	122015
860	2016	Dec	4.7	12	122016
861	2017	Dec	4.1	12	122017
862	2018	Dec	3.9	12	122018
863	2019	Dec	3.5	12	122019

Transform

Fuel DataFrame

- We pulled Month and Year into separate columns
- Created a Month/Year Column to be able to group by off both.
- Reduced columns:
  - Date
  - Month/Year
  - Regular, Premium and Diesel
- Sorted by date

	monthyear	date	regular	premium	diesel
0	11995	1995-01-30	1.08180	1.27480	1.09820
27	21995	1995-02-27	1.07250	1.26475	1.08775
53	31995	1995-03-27	1.07200	1.25800	1.08750
79	41995	1995-04-24	1.11125	1.29475	1.10400
105	51995	1995-05-29	1.17840	1.36420	1.12500
...	...	...	...	...	...
234	92020	2020-09-28	2.18275	2.84750	2.41375
260	102020	2020-10-26	2.15800	2.82175	2.38875
286	112020	2020-11-30	2.10820	2.78220	2.43200
312	122020	2020-12-28	2.19525	2.85025	2.58475
26	12021	2021-01-25	2.33425	2.97525	2.68050

Load

- Created corresponding tables in Database in PgAdmin.
- Loaded data from jupyter notebook.
- Performed a few example queries to demonstrate database capabilities:
  - Regular > $3.80
  - Unemployment Rate > 9.5

	regular real	diesel real	rate real	month character varying (3)	year integer
1	3.85175	4.1265	8.2	Mar	2012
2	3.9004	4.115	8.2	Apr	2012
3	3.9062	4.0468	9	May	2011
4	4.0542	4.6768	5.6	Jun	2008
5	4.0615	4.703	5.8	Jul	2008
6	3.8485	4.12	7.8	Sep	2012

	regular real	diesel real	rate real	month character varying (3)	year integer
1	2.715	2.84475	9.8	Jan	2010
2	2.644	2.7845	9.8	Feb	2010
3	2.7716	2.9148	9.9	Mar	2010
4	2.84825	3.059	9.9	Apr	2010
5	2.8362	3.0688	9.6	May	2010
6	2.6164	2.6338	9.6	Aug	2009
7	2.554	2.626	9.8	Sep	2009
8	2.55125	2.672	10	Oct	2009
9	2.6514	2.7922	9.9	Nov	2009
10	2.859	3.14	9.8	Nov	2010
11	2.60725	2.7445	9.9	Dec	2009

Questions?