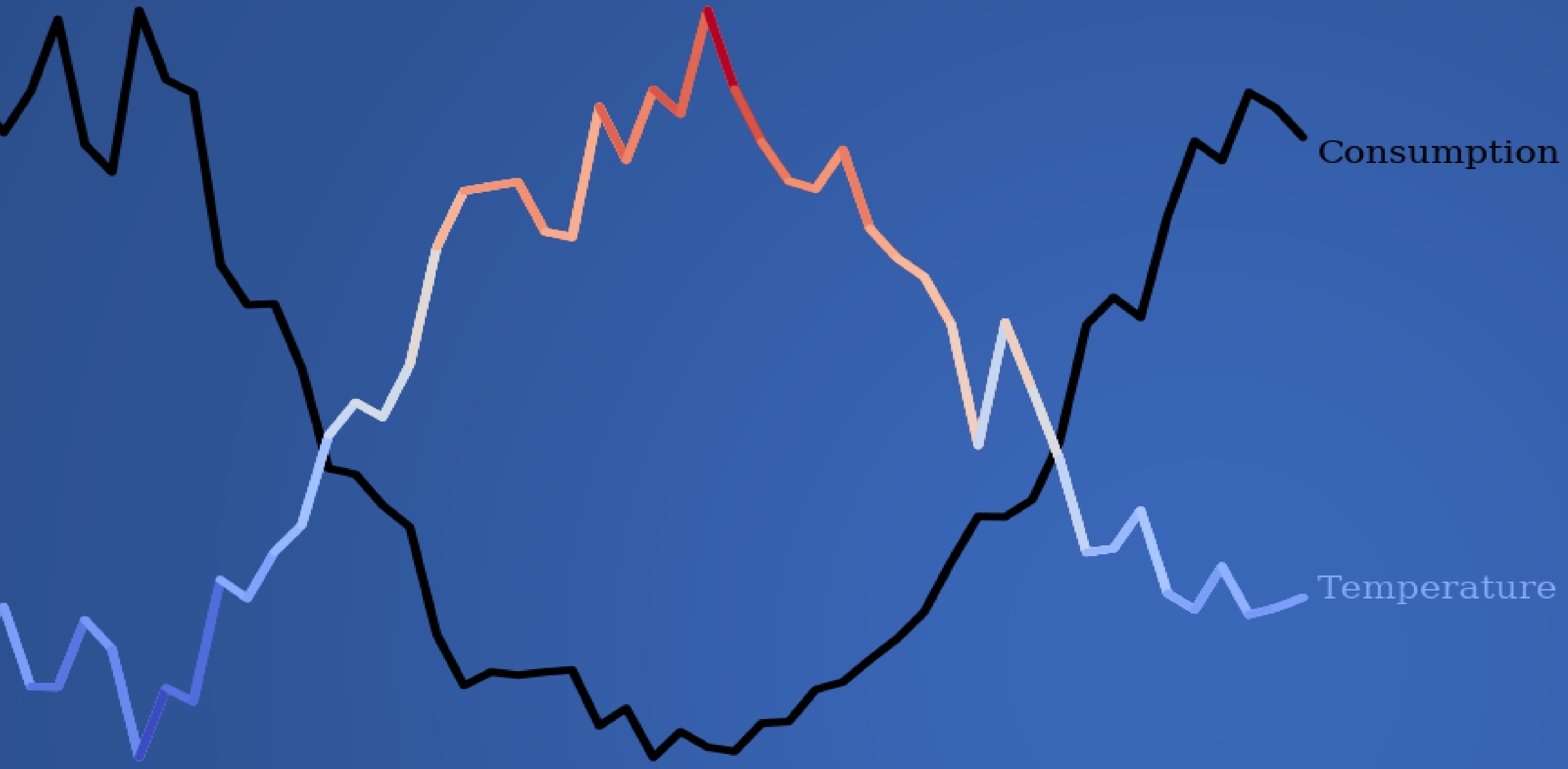


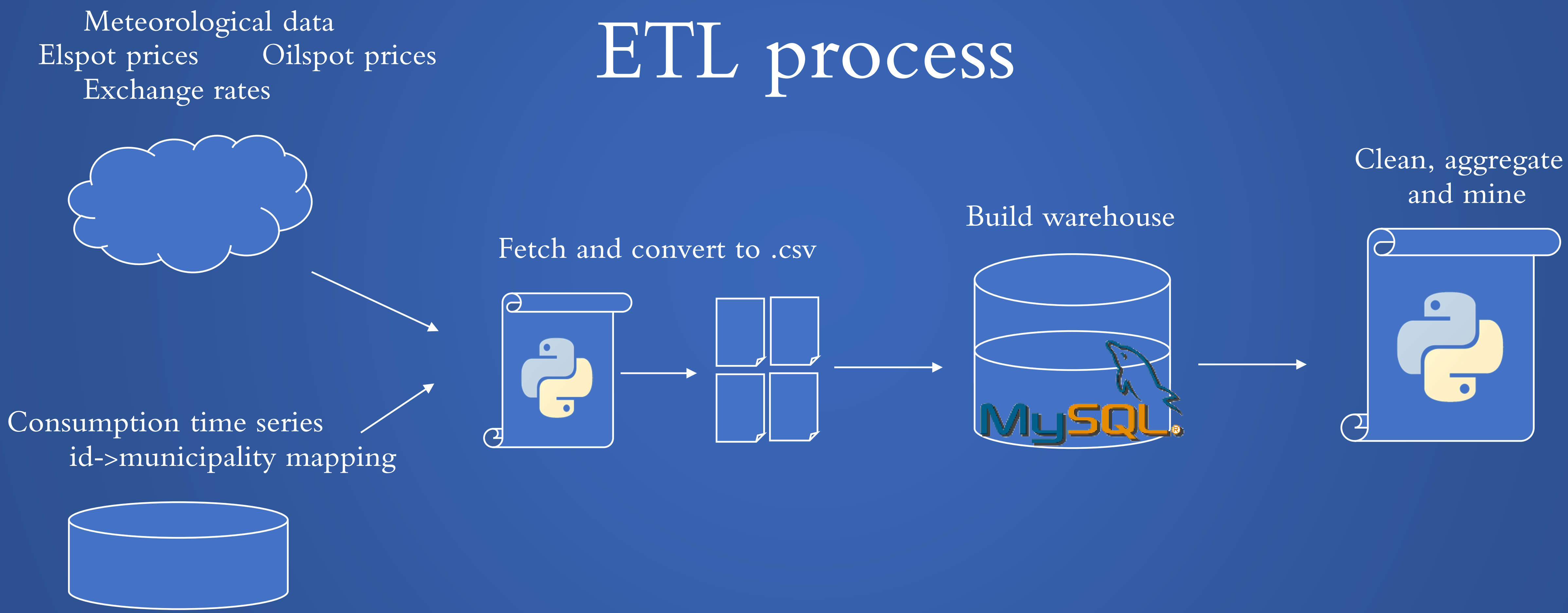
The intermingling of European oil prices, elspot prices, local weather, and electrical energy consumption in Agder, Norway

Arild Madshaven | arild.madshaven@epfl.ch

Motivation

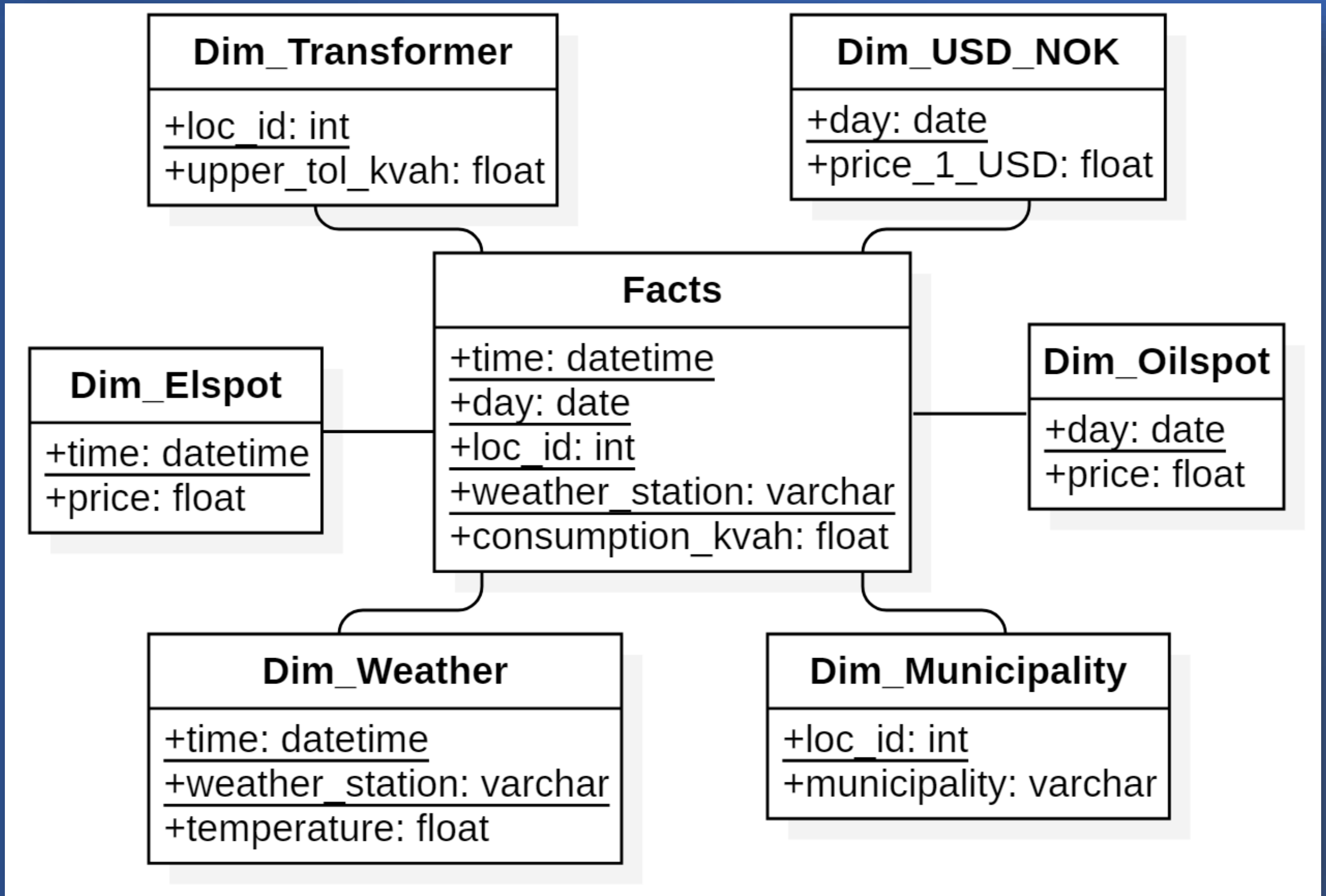
- Smart householdings and new smartphone apps might give ‘smart customers’ the option to regulate their consumption based on el prices. Do they already exist, and can they be identified?
- Do customers oppose to rapidly changing prices?
- Upgrading the electrical distribution network is costly, which can incentivize energy providers to negotiate with atypical consumers in order to limit transformer loads
- 98 % of electrical power production in Norway is renewable. How much do oil prices affect the elspot prices in Norway?

ETL process


Meteorological data
Elspot prices
Oilspot prices
Exchange rates
Consumption time series
id->municipality mapping
Fetch and convert to .csv
Build warehouse
MySQL
Clean, aggregate
and mine

Warehouse structure

- Star schema
- Sample query: “Report transformer overloads”

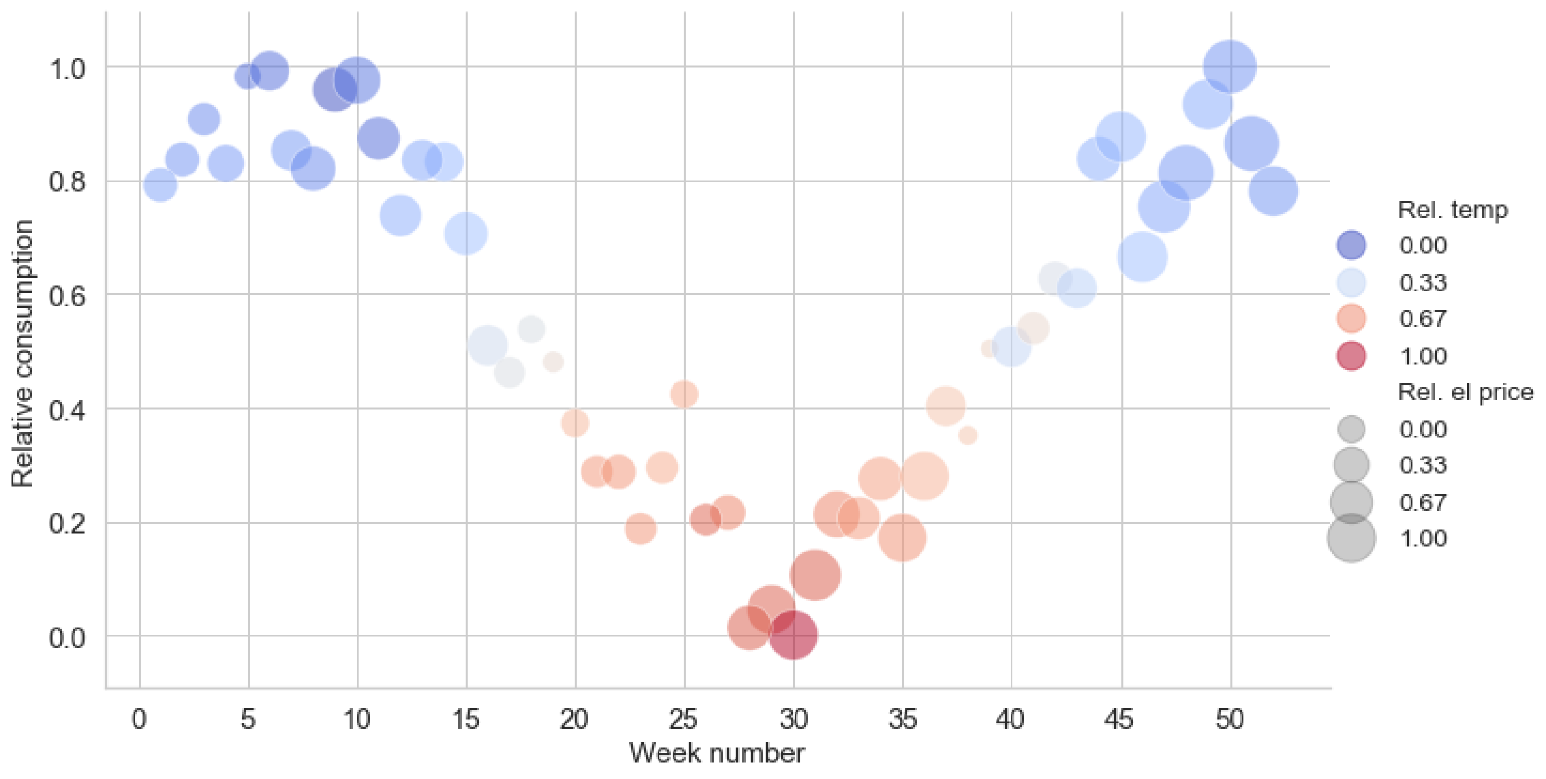



```
query = """
    select
        f.loc_id,
        time,
        consumption_kvah,
        upper_tol_kvah
    from Facts f left join Dim_Transformer t on f.loc_id = t.loc_id
    where consumption_kvah > upper_tol_kvah
"""
pd.read_sql(query, conn)
```

	loc_id	time	consumption_kvah	upper_tol_kvah
0	2	2018-12-12 08:00:00	118.651801503391	100
1	2	2018-12-24 14:00:00	106.169675519896	100
2	2	2018-12-24 16:00:00	102.122475488993	100
3	2	2018-12-24 17:00:00	104.576527002956	100
4	2	2018-12-24 18:00:00	106.617306287488	100
5	2	2019-01-27 21:00:00	100.62430123981	100
6	2	2019-10-27 02:00:00	107.557775823303	100

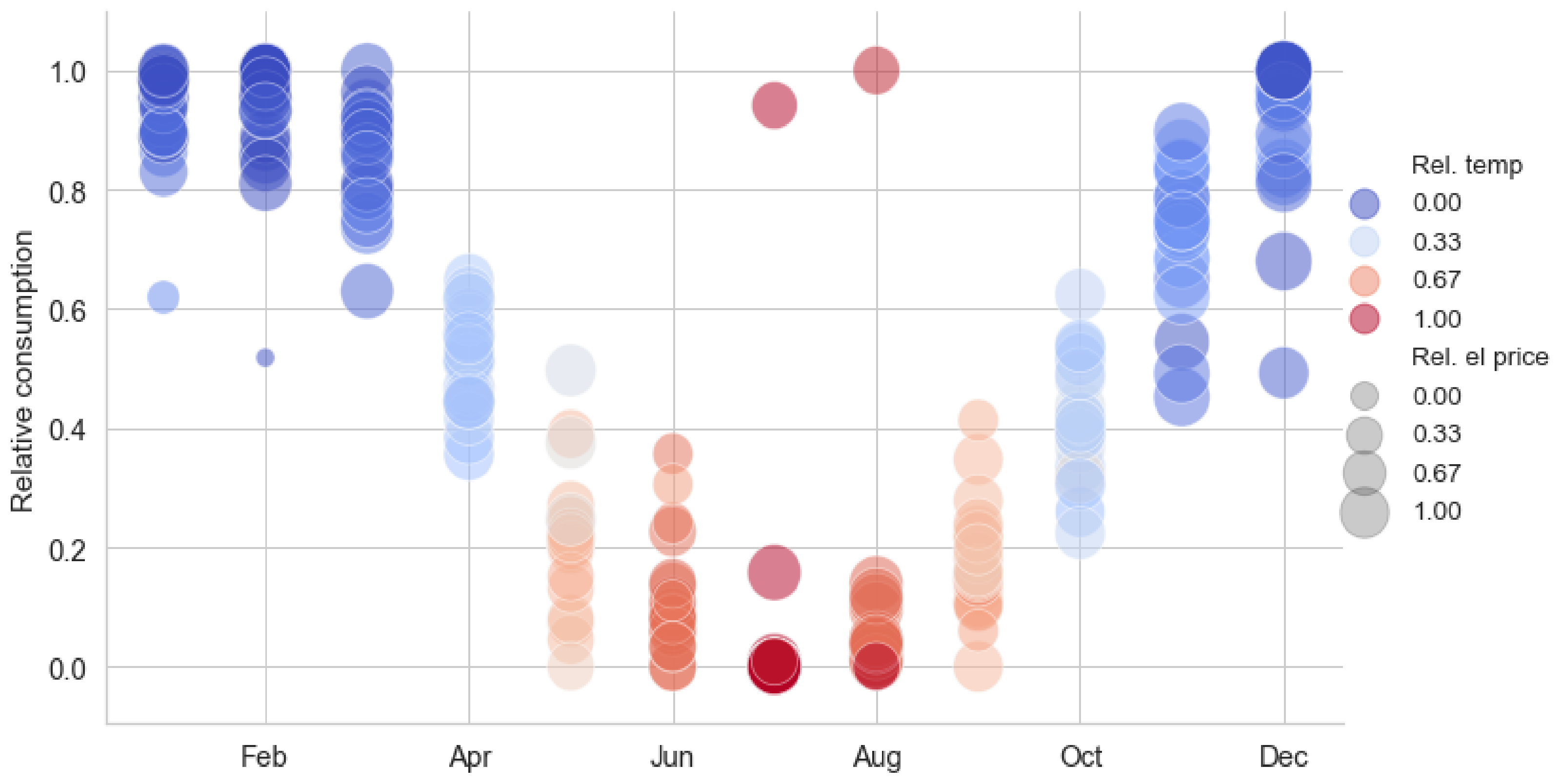
Data insight

- Aggregated consumption per week for one sample customer
- Take home:
  - it's hotter during summer than winter
  - consumption peaks ~ first and last 10 weeks of the year
  - prices are typically high in Nov/Dec
  - most customers follow this trend



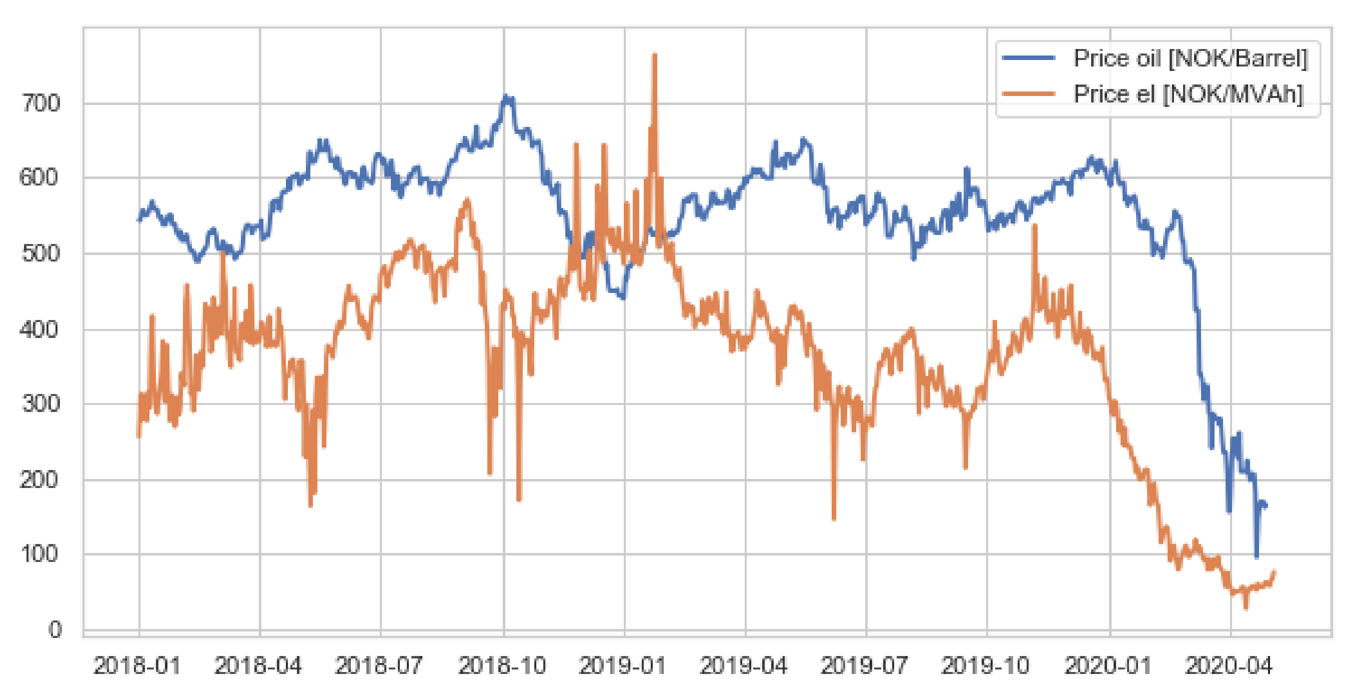
Data insight

- Aggregated consumption per month for all customers
- Notice a few outliers during summer. These might belong to special commercial customers, or perhaps a cabin mostly used during summer
- General strong trend towards high consumptions during winter for most customers



Data insight

- Oil- and elspot prices had a dramatic start to 2020
- There are several reasons for this, many of which are not included in the model. Results related to oil and el prices co-development should be considered with caution due to the simple nature of the data foundation
- Factors such as wind, down-pour, and variations in the European marked certainly affect prices, despite their absence in this project



Key algorithms

- The Granger causality test
  - Given two time varying signals, X, Y, suppose we want to forecast X_{t+1} using past values of X
  - Suppose forecasting X_{t+1} is more successful by including past values of Y
  - Then Y contains information about X that is not found elsewhere (specifically not in past values of X itself)
  - Assuming the series of Y_t used to forecast X_{t+1} takes place before X_{t+1} in time, Y is said to Granger cause X if it contributes to a more successful forecasting of X_{t+1}
  - In practice carried out by a series of F-tests on lagged values

Key findings

- From Granger causality tests
  - Notice: Null hypothesis is “column X does NOT cause row Y”. A p-value lower than 0.05 means we reject the null hypothesis
- Customers differ. Trends illustrated here have been picked up using weekly aggregates and maximum 5 weeks of memory
- **Customers 3, 4, 11, 16,** and **18** seem to adapt consumption to temperatures, and **customers 8, 10,** and **14** seem to adapt to elspot prices
- Results vary slightly based on resolution (monthly, daily) but trends are fairly consistent

P-values for customer 10

	consumption_kvah	temperature	el_price
consumption_kvah	1.0000	0.1451	0.0452

P-values for customer 11

	consumption_kvah	temperature	el_price
consumption_kvah	1.0000	0.0262	0.3244

- Strong indications that temperatures regulate elspot prices (p=0.0022) using monthly resampling and 5 months of memory. This trend is strong regardless of resolution
- Surprising suggestion: elspot prices regulate oilspot prices and not the other way around
- Keep in mind the development of el and oil prices as of late
- Oil price drop during winter might have strengthened the idea of an oil->temperature relationship (p=0.0762). As we only analyze two years of data, I consider this a plausible theory

	temperature	price_oil	price_el
temperature	1.0000	0.0762	0.4581
price_oil	0.2361	1.0000	0.0263
price_el	0.0022	0.1726	1.0000

The **apriori algorithm for frequent itemset mining**

- Mines for frequent itemsets and predictive rules
- Does NOT equal causality, nor correlation. A rule on the form BODY->HEAD is interpreted as follows: If BODY, then probably also HEAD
- Support: How common is the itemset? $\frac{\#\ transactions\ containing\ BODY\ and\ HEAD}{total\ \#\ of\ transactions}$
- Confidence: When BODY, how often HEAD? $\frac{\#\ transactions\ containing\ BODY\ and\ HEAD}{\#\ transactions\ containing\ BODY}$
- Lift: How interesting is the rule? Is HEAD popular regardless of BODY? Lift = $\frac{confidence}{support(HEAD)}$. We're looking for rules with lift > 1

```
transactions = [
    ['beer', 'diapers'],
    ['diapers', 'milk'],
    ['beer', 'nuts', 'diapers'],
    ['milk', 'nuts']
]
```

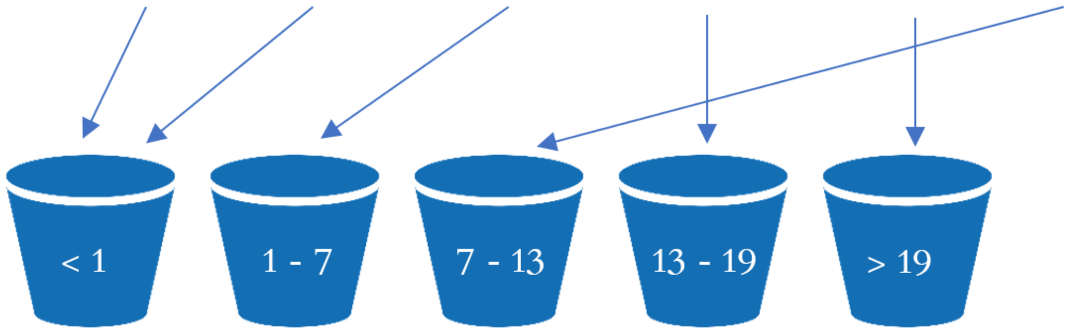
```
Items:  {'beer', 'diapers'}
Support: 0.5000
BODY -> HEAD[Confidence, Lift]

        {'beer'}  ->  {'diapers'}[1.00, 1.33]
```

The problem of data bagging

- Since we're dealing with numerical data(and not categorical such as "diapers", "beer", the result is **heavily dependent** on how you **bag** and aggregate the data
- Two 'bucketing' approaches:
  - **Based on value: Number of standard deviations from mean**
    - Prone to noise (outliers always considered)
  - **Based on dynamics: N buckets from min to max**
    - Number of buckets is crucial
    - More buckets limits the impact of unlucky thresholding, but reduces generalizability

Week no.	1	2	3	4	5	6
Temperature	-5	0	5	15	25	10



Week no.	1	2	3	4	5	6
Bucket diff	NaN	0	+1	+2	+1	-2
diff %	NaN	0 %	+20 %	+40 %	+20 %	-40 %

Key findings

- **On temperature and el prices (with median hourly temperatures):**
  - **Extremely high el prices -> Low temperatures** [sup 93 rows, conf 0.87, lift 5.26]
    > 99.8 percentile (2.3, 15.9) percentile
  - **Extremely low el prices -> Normal temperatures** [sup 66 rows, conf 0.96, lift 1.50]
    < 0.2 percentile (34.1, 68.2) percentile
  - **Big drop in el prices -> Jump in temperatures** [sup 18 months, conf 0.95, lift 3.84]
    down 40% from last month up 10% from last month (aggregated months)
  - **Big drop in el prices -> Small jump in temperatures** [sup 18 months, conf 0.95, lift 3.84]
    down 40% from last month up 10% from last month (aggregated months)
  - **Big jump in temperatures -> Drop in el prices** [sup 12 months, conf 0.86, lift 5.28]
    up 40% from last month down 20% from last week (aggregated months)

- On consumption and el prices:
  - Big jump in el prices -> Drop in consumption [sup 12 months, conf 0.75, lift 3.84]
    up 30% from last month down 10% from last month (aggregated months)
  - Apart from this rule, itemsets containing el prices and consumption also contain temperatures, meaning most shifts in consumption also cooccur with shifting temperatures, and not only shifting prices. Strong examples include:
    - {Slight drop in consumption, small jump in el prices} -> slightly increasing temperatures
    - {Big drop in consumption, big drop in el prices} -> slightly increasing temperatures
    - {Big drop in consumption, big jump in temperatures} -> decreasing el prices
- In general, both el prices and consumption seem to be heavily linked to shifting temperatures, and not so much to one another
- On oil and el prices:
  - Extremely low oil prices -> Very low el prices [sup 35 days, conf 1.00, lift 10.96]
    < 0.2 percentile (0.2, 2.3) percentile (aggregated days)

Summary

- **General data insight** reveals that, for most customers, consumption decreases when temperatures drop
- **The apriori algorithm** reveals that, in general, prices are high when temperatures are low, and vice versa
- Of most commercial value might be the results from the **Granger causality tests**, which reveal the individual differences between customers, and their concerns with prices and temperatures
- Information about customer differences might ease the problem of capacity planning. Especially customers with abnormal consumption time series might be of interest to the el providers
- Suggestions that el prices cause oil prices. This relationship is definitely more intricate than can be encapsulated by this project

Special thanks

To Agder Energi Nett AS for consumption time series donation
https://www.ae.no/en/

Source code for this project can be found at
https://github.com/arilmad/BI_project

and I can personally be reached by email
arild.madshaven@epfl.ch OR (preferably) arilmad@stud.ntnu.no