

fiRst Group Project

Ismail Batur Usta / Efehan Danisman / Ozgur Ozdemir

What we did?

- ▶ We analyzed cuts at the power plants in Turkey between 2012-2018.
- ▶ We had in total 73313 observations with 8 variables
- ▶ We mutated new observations from the existing ones: Plant.Type, Duration of Cut, Capacity Ratio at the cut and reason of the cut.
- ▶ We tidied the raw data using regular expressions and stringr package.
- ▶ We used tidy text mining to analyze count of words and which word is following which word.
- ▶ We divided cuts into two category, Malfunctions and Planned Activities and looked for their distributions.
- ▶ We looked at differences between malfunctions and planned activities in terms of duration of the cut.
- ▶ We looked at malfunction types, malfunction reasons and durations according to plant type.

Cuts At Power Plants in Turkey(2012-2018)

Yearly Incidents are way higher at 2018.

```
## Error in ggplot(data = yearly_cuts, aes(y = Start.Date,
```

Glimpse of Cleaning

It was not easy

```
cutsPlant.Name <- cutsPlant.Name %>% str_replace_all("[ı]",  
"i") %>% str_replace_all("enerj.sa", "enerjisa") %>%  
str_replace_all("yenikoy ts", "yenikoy tes") %>%  
str_replace_all("ienikoi tes", "yenikoy tes") %>%  
str_replace_all("^ova elektrik", "gebze ova elektrik") %>%  
str_replace_all("yatagan .", "yatagan tes") %>%  
str_replace_all("kokluce$", "kokluce hes") %>% str_replace_all(".  
entek", "entek") %>% str_replace_all("kurtun-hes", "kurtun hes")  
%>% str_replace_all("^rwe_turcas_guney", "denizli  
rwe_turcas_guney") %>% str_replace_all("tekirdag  
santrali.", "modern enerji tekirdag santrali") %>%  
str_replace_all("karadag", "karadag res") %>%  
str_replace_all(".?menzelet( hes)ı", "menzelet hes") %>%  
str_replace_all("\\.", "") %>% str_replace_all("hidro(\\s?elektrik  
santral[ıı]) e\\.?s)", "hes") %>% str_replace_all("(termik  
santral[ıı])\\s?sts\\s?)", "tes") %>%  
str_replace_all("tuncbilektes", "tuncbilek tes") %>%
```

Overview of Plant Categories

We've categorised power plants by their type, doing analysis by plant name would not yield much useful results.

- *HES: Hydroelectricity Plant

- *TES: Thermal Energy Plant

- *RES: Wind Energy Plant(Wind Turbines)

- *DGKC: Natural Gas Combined Cycle Plant

- *JES: Geothermal Energy Plant

Overview of Plant Categories-cont'd.

```
## Error in eval(lhs, parent, parent): 'cuts' nesnesi bulur
```

Cut Reason by Text Mining

```
## Error in ggplot(bigram_counttop20, aes(reorder(bigram, r
## Error in eval(lhs, parent, parent): 'cuts' nesnesi bulur
```

Cut Reason by Plant

```
## Error in eval(lhs, parent, parent): 'catmalf' nesnesi bu
```

```
## Error in labels %||% p$x$attrs[[1]][["labels"]]: 'm_by_t
```


Shutdown Reason by Category

```
## Error in eval(lhs, parent, parent): 'catmalf' nesnesi bu
```

Cut Reason by Category

```
## Error in eval(lhs, parent, parent): 'cuts' nesnesi bulur
```

```
## Error in as.vector(m_count$Plant.Name): 'm_count' nesnes
```

```
## Error in eval(lhs, parent, parent): 'cuts' nesnesi bulur
```

```
## Error in eval(expr, envir, enclos): 'malf' nesnesi bulur
```

```
## Error in ggplot(malf, aes(x = quarter)): 'malf' nesnesi
```

Conclusions

- ▶ Most time consuming part was data transformation and cleaning.
- ▶ Especially in 2018, number of data entries have drastically increased.
- ▶ While in average Thermal plants produce higher amounts of power, on total Hydroelectric plants' throughput is the highest.
- ▶ Each type of plant have a different leading reason for shutdowns.

Thanks

```
## Error in inherits(x, "Source"): 'cuts' nesnesi bulunamadı
```

```
## Error in tm_map(cutsReason.Corpus, PlainTextDocument):
```

```
## Error in tm_map(cutsReason.Corpus, tolower): 'cutsReason'
```

```
## Error in wordcloud(cutsReason.Corpus, min.freq = 5, max.:
```