Polars cheat sheet

Open in Colab

General

Install

```
pip install polars
```

Import

```
import polars as pl
```

Creating/reading DataFrames

```
# Create DataFrame
df = pl.DataFrame(
  {
    "nrs": [1, 2, 3, None, 5],
    "names": ["foo", "ham", "spam", "egg", None
    "random": [0.3, 0.7, 0.1, 0.9, 0.6],
    "groups": ["A", "A", "B", "C", "B"],
  }
)

# Read CSV
df = pl.read_csv("https://j.mp/iriscsv",
                 has_header=True)

# Read parquet
df = pl.read_parquet("path.parquet")
```

Subset Observations - rows

```
# Filter: Extract rows that meet logical criter
df.filter(pl.col("random") > 0.5)
df.filter(
  (pl.col("groups") == "B")
  & (pl.col("random") > 0.5)
)


# Sample
# Randomly select fraction of rows.
df.sample(frac=0.5)

# Randomly select n rows.
df.sample(n=2)


# Select first n rows
df.head(n=2)

# Select last n rows.
df.tail(n=2)
```

Subset Variables - columns

```
# Select multiple columns with specific names.
df.select(["nrs", "names"])


# Select columns whose name matches regex.
df.select(pl.col("^n.*$"))
```

by Franz Diebold (diebold.io)