

Python Polars Commands

Dr. Lema LOGAMOU SEKNEWNA

```
# Import & Create DataFrame
```

```
## import polars as pl
```

```
df = pl.DataFrame({  
    "name": ["Alice", "Bob"],  
    "age": [25, 32]  
})
```

```
df = pl.read_csv("data.csv")
```

```
df = pl.read_parquet("data.parquet")
```

View data

```
df.head()
```

```
df.tail(3)
```

```
df.sample(5)
```

```
df.describe()
```

```
df.shape
```

```
df.columns
```

```
df.dtypes
```

Select & filter

```
df.select("name")
```

```
df.select(["name", "age"])
```

```
df.filter(pl.col("age") > 30)
```

```
df.with_columns(  
    (pl.col("salary") * 1.1).alias("new_salary")  
)
```

```
df.drop("salary")
```

GroupBy & Aggregate

```
df.groupby("name").agg(  
    pl.col("salary").sum().alias("total_salary"),  
    pl.col("age").mean().alias("avg_age")  
)
```

Sort & Unique

```
df.sort("age")
```

```
df.sort("age", descending=True)
```

```
df.unique()
```

```
df.unique(subset=["name"])
```

Joining data

```
df1.join(df2, on="id", how="inner")
```

```
df1.join(df2, on="id", how="left")
```

Lazy mode

```
q = (  
    pl.scan_csv("big.csv")  
    .filter(pl.col("age") > 30)  
    .groupby("gender")  
    .agg(pl.col("salary").mean())  
)
```

```
result = q.collect()
```



```
# Write data
```

```
df.write_csv("output.csv")
```

```
df.write_parquet("output.parquet")
```

Pandas interop

```
pandas_df = df.to_pandas()  
df = pl.from_pandas(pandas_df)
```

More advanced tips

Use **LazyFrame** for big data

pl.col() for expressions

Use **.pipe()** for custom functions

Combine **.with_columns()** & **.select()**