Exploring brainlat dataset

```
import polars as pl
import seaborn.objects as so
from tabulate import tabulate

root = "/Volumes/T7/BrainLat/"
records_path = f"{root}/BrainLat_records.csv"
only_eeg = pl.scan_csv(records_path).filter(pl.col("eeg") == 1.0).collect()
display(only_eeg)
```

```
counts = only_eeg.select(["diagnosis", "eeg"]).group_by(
    "diagnosis").agg(pl.all().sum())

cognition_path = f"{root}/Demographic and cognition/BrainLat_Cognition.csv"
cognition = (
    pl.scan_csv(cognition_path)
    .filter((pl.col("id").is_in(only_eeg.get_column("id").to_list())))
    .collect()
)
demographics_path = (
    "/Volumes/T7/BrainLat/Demographic and cognition/BrainLat_Demographic.csv"
)
demographics = (
    pl.scan_csv(demographics_path, dtypes={"Age": float})
    .filter((pl.col("id").is_in(only_eeg.get_column("id").to_list())))
    .collect()
    .with_columns(pl.col("Age").cast(pl.Int64, strict=False))
)
display(counts)
display(demographics)
display(cognition)
```

```
cognition_summary = cognition.group_by("diagnosis").agg(pl.all().count())
display(cognition_summary)
```

```
df = cognition_summary.to_pandas().T
df.columns = df.iloc[0]
df = df.iloc[1:]
print(tabulate(df, headers="keys", tablefmt="grid"))
```

```
+------------------------------+------+-------+------+------+------+
|                              |   AD |   FTD |   CN |   PD |   MS |
+==============================+======+=======+======+======+======+
| id                           |   35 |    16 |   42 |   27 |   32 |
+------------------------------+------+-------+------+------+------+
| moca_total                   |   29 |    14 |   30 |   26 |    0 |
+------------------------------+------+-------+------+------+------+
| moca_visuospatial            |   29 |    14 |   30 |   19 |    0 |
+------------------------------+------+-------+------+------+------+
| moca_recog                   |   29 |    14 |   30 |   19 |    0 |
+------------------------------+------+-------+------+------+------+
| moca_attention               |   29 |    14 |   30 |   26 |    0 |
+------------------------------+------+-------+------+------+------+
| moca_language                |   29 |    14 |   30 |   26 |    0 |
+------------------------------+------+-------+------+------+------+
| moca_abstraction             |   29 |    14 |   30 |   26 |    0 |
+------------------------------+------+-------+------+------+------+
| moca_memory                  |   29 |    14 |   30 |   26 |    0 |
+------------------------------+------+-------+------+------+------+
| moca_orientation             |   29 |    14 |   30 |   26 |    0 |
+------------------------------+------+-------+------+------+------+
| ifs_total_score              |   33 |    14 |   27 |   25 |   28 |
+------------------------------+------+-------+------+------+------+
| ifs_motor_series             |   31 |    14 |   24 |   17 |   28 |
+------------------------------+------+-------+------+------+------+
| ifs_conflicting_instructions |   30 |    14 |   24 |   17 |    0 |
+------------------------------+------+-------+------+------+------+
| ifs_motor_inhibition         |   31 |    14 |   23 |   16 |    0 |
+------------------------------+------+-------+------+------+------+
| ifs_digits                   |   31 |    14 |   23 |    7 |    0 |
+------------------------------+------+-------+------+------+------+
| ifs_months                   |   31 |    13 |   23 |    0 |    0 |
+------------------------------+------+-------+------+------+------+
```

ifs_visual_wm	31	14	23	24	0
ifs_proverb	31	14	23	7	0
ifs_verbal_inhibition	31	16	25	0	0
mini_sea_fer	31	16	29	25	0
mini_sea_tom	31	16	29	24	0
emotion recog	0	0	0	0	0

First lets have 2 models first model that is used only on behavioral variables, then the second we use only on the people who have a eeg data

```
cognition_path = "/Volumes/T7/BrainLat/Demographic and cognition/BrainLat_Cognition.csv"
cognition_all = (
    pl.scan_csv(cognition_path)
    .collect()
    .filter(~pl.all_horizontal(pl.all().is_null()))
    .group_by("diagnosis")
    .agg(pl.all().count())
)
display(cognition_all)
```

```
df_all = cognition_all.to_pandas().T
df_all.columns = df_all.iloc[0]
df_all = df_all.iloc[1:]
print(tabulate(df_all, headers="keys", tablefmt="grid"))
```

	FTD	PD	AD	MS	CN
id	165	56	279	34	250
moca_total	94	55	111	0	156
moca_visuospatial	95	20	111	0	154
moca_recog	95	20	111	0	154

moca_attention	95	33	111	0	154
moca_language	95	33	111	0	154
moca_abstraction	95	33	111	0	154
moca_memory	95	33	111	0	154
moca_orientation	95	33	111	0	154
ifs_total_score	115	54	177	29	166
ifs_motor_series	112	18	175	29	162
ifs_conflicting_instructions	112	18	174	0	162
ifs_motor_inhibition	112	17	175	0	161
ifs_digits	112	13	176	0	161
ifs_months	112	0	176	0	161
ifs_visual_wm	113	53	176	0	161
ifs_proverb	113	13	176	0	161
ifs_verbal_inhibition	119	0	182	0	169
mini_sea_fer	59	32	169	0	113
mini_sea_tom	59	31	162	0	122
emotion recog	0	22	0	0	0