**Spark SQL**

// Most common words in job title

df.select(explode(split(lower(col("title")), " ")).alias("word"))

.groupBy("word")

.agg(count("\*").alias("count"))

.filter(col("count") > 5)

.orderBy(col("count").desc())

// Most recent jobs

df.orderBy(col("date").desc()).limit(10)

// Companies with the most job postings

df.groupBy("company")

.agg(

count("\*").alias("jobCount"),

collect\_set("title").alias("titles") )

.orderBy(col("jobCount").desc())

.limit(10)

// Jobs with salary info

df.filter((col("salary").isNotNull()) & (col("salary") != ""))

// Rank job hiring organization by number of job postings

df.groupBy("company")

.agg(count("\*").alias("job\_count"))

.orderBy(col("job\_count").desc()).limit(10)

// Job postings with the word data

df.filter(lower(col("title")).contains("data"))

.count()

// Number of job postings by state

df.groupBy("location")

.agg(count("\*").alias("job\_count"))

.orderBy(col("job\_count").desc()