

```
#Install & Set Up PySpark

!pip install pyspark

#Import Required Libraries

from pyspark.sql import SparkSession from pyspark.sql.functions import to_date, month,
year, sum as _sum

# Create Spark Session

spark = SparkSession.builder.appName("MonthlyRevenue").getOrCreate()

# Load CSV File (Upload your CSV file first in Colab)

from google.colab import files uploaded = files.upload()

#Replace with your file name (change if needed)

csv_file = "revenue_data.csv"

# Read Data

df = spark.read.csv(csv_file, header=True, inferSchema=True)

# Convert Date Column to Proper Date Format

df = df.withColumn("Date", to_date(df["Date"], "yyyy-MM-dd"))

# Extract Year and Month

df = df.withColumn("Year", year("Date")).withColumn("Month", month("Date"))

# Group by Year and Month, Calculate Revenue

monthly_revenue = df.groupBy("Year",
"Month").agg(_sum("Revenue").alias("Total_Revenue"))

# Show Results

monthly_revenue.orderBy("Year", "Month").show()
```