

```
from pyspark.sql import SparkSession
from pyspark.sql.functions import col, lag, concat_ws, coalesce
from pyspark.sql.window import Window

# Initialize Spark session
spark = SparkSession.builder.appName("WebTreeExample").getOrCreate()
# 1) Read the Input_data CSV file into a PySpark DataFrame

input_df =
spark.read.csv('/Users/ishayadav/PycharmProjects/PythonProject/Input_data.csv',
header=True, inferSchema=True, quote='')

windowSpec = Window.orderBy("DVISION_ID", "CLASS_ID", "BRAND_ID",
"PARENT_CATEGORY_ID")

# Add the Web_TREE column
df_out = input_df.withColumn(
    "Web_TREE",
    concat_ws(
        "_",
        coalesce(lag("PARENT_CATEGORY_ID").over(windowSpec),
col("PARENT_CATEGORY_ID")),
        col("PARENT_CATEGORY_ID")
    )
)
df_out.show()
# Write the DataFrame as a JSON file
df_out.write.json("/Users/ishayadav/PycharmProjects/PythonProject/Downloads/assessment3./out/json")

# Write the DataFrame as a Parquet file
df_out.write.parquet("/Users/ishayadav/PycharmProjects/PythonProject/Downloads/assessment3./out/parquet")
```