Lab 07

Ivan Villasmil

November 2, 2025

1 Part 1: Load the Dataset

```
# Loading required libraries
import numpy as np
import pandas as pd
import plotly.express as px
import plotly.io as pio
import plotly.graph_objects as go
from pyspark.sql import SparkSession
from pyspark.sql.functions import (
    col, split, explode, regexp_replace, transform, when,
    to_date, monotonically_increasing_id, expr
)
from pyspark.sql import functions as F
import re

np.random.seed(950)
pio.renderers.default = "notebook"

# Initializing Spark Session
spark = SparkSession.builder.appName("LightcastData").getOrCreate()

# Uploading CSV file into a Spark DataFrame
df = spark.read.option("header", "true").option("inferSchema", "true").option("multiLine","t

# Verifying Data: Display Schema (column names & data types)
# print("---This is Diagnostic check, No need to print it in the final doc---") # Comment li
# df.printSchema() # Comment line when rendering the submission
```

```
# Verifying Data: Display first five rows
# print("---This is Diagnostic check, No need to print it in the final doc---") # Comment lin
df.show(20)  # Comment line when rendering the submission

# Typecasting numeric columns to double
df_cleaned = df.withColumn("SALARY", col("SALARY").cast("double"))
df_cleaned = df.withColumn("SALARY_FROM", col("SALARY_FROM").cast("double"))
df_cleaned = df.withColumn("SALARY_TO", col("SALARY_TO").cast("double"))
df_cleaned = df.withColumn("MIN_YEARS_EXPERIENCE", col("MIN_YEARS_EXPERIENCE").cast("double"))
df_cleaned = df.withColumn("MAX_YEARS_EXPERIENCE", col("MAX_YEARS_EXPERIENCE").cast("double"))
df_cleaned = df.withColumn("DURATION", col("DURATION").cast("double"))
df_cleaned = df.withColumn("MODELED_DURATION", col("MODELED_DURATION").cast("double"))

# Typecasting date columns to M/d/yyyy
df_cleaned = df.withColumn("POSTED", to_date(col("POSTED"), "M/d/yyyy"))
df_cleaned = df.withColumn("EXPIRED", to_date(col("EXPIRED"), "M/d/yyyy"))
df_cleaned = df.withColumn("LAST_UPDATED_DATE", to_date(col("LAST_UPDATED_DATE"), "M/d/yyyy"))
df_cleaned = df.withColumn("MODELED_EXPIRED", to_date(col("MODELED_EXPIRED"), "M/d/yyyy"))

# Verifying schema
df_cleaned.select(
    "SALARY", "SALARY_FROM", "SALARY_TO",
    "MIN_YEARS_EXPERIENCE", "MAX_YEARS_EXPERIENCE",
    "DURATION", "MODELED_DURATION",
    "POSTED", "EXPIRED", "LAST_UPDATED_DATE", "MODELED_EXPIRED"
).printSchema()

# Convert Spark DataFrame to Pandas for final cleaning
# pdf_cleaned = df_cleaned.toPandas()

# Fill missing values with appropriate defaults
# pdf_cleaned["SALARY"].fillna(0, inplace=True)
# pdf_cleaned["SALARY_FROM"].fillna(0, inplace=True)
# pdf_cleaned["SALARY_TO"].fillna(0, inplace=True)
# pdf_cleaned["MIN_YEARS_EXPERIENCE"].fillna(0, inplace=True)
# pdf_cleaned["MAX_YEARS_EXPERIENCE"].fillna(0, inplace=True)
# pdf_cleaned["DURATION"].fillna(0, inplace=True)
# pdf_cleaned["MODELED_DURATION"].fillna(0, inplace=True)

# Optional: fill missing dates with a placeholder or drop them
# pdf_cleaned["POSTED"].fillna(pd.Timestamp("1900-01-01"), inplace=True)
# pdf_cleaned["EXPIRED"].fillna(pd.Timestamp("1900-01-01"), inplace=True)
```

```
# pdf_cleaned["LAST_UPDATED_DATE"].fillna(pd.Timestamp("1900-01-01"), inplace=True)
# pdf_cleaned["MODELED_EXPIRED"].fillna(pd.Timestamp("1900-01-01"), inplace=True)

# Save cleaned data to CSV
# pdf_cleaned.to_csv("data/lightcast_cleaned.csv", index=False)
```

```
[Stage 25:>                                                         (0 + 1) / 1]
```

```
+--------------------+-----------------+----------------------+----------
+--------+---------+--------+--------------------+--------------------+--------------------
+-----------+--------------------+--------------------+--------------------
+---------------+----------------+---------+--------------------+--------------------
+-------------------+--------------------+---------------------+-------------
+--------------------+-------------+-------------------+---------------+--------------------
+--------------------+--------------------+-------------+------+-----------
+-----------------+-------------------+---------+-----------+--------------------
+--------------------+------------------+------+--------------------+-----
+--------------------+-----+-------------+---------------+--------------------
+---------------+--------------------+------------+--------------------
+------------+--------------------+------+--------------------+------+--------------------
+------+--------------------+------+--------------------+------+--------------------
+------------------+--------------------+--------------------+--------------------
+--------------------+--------------------+----------------------+--------------------
+--------------------+--------------------+--------------------+--------------------
+--------------------+----------+--------------------+----------+--------------------
+--------------------+--------------------+--------------------+--------------------
+--------------------+--------------------+----------+--------------------
+----------+--------------------+----------+---------------+----------+---------------
+---------------+-------------------+--------------+--------------------
+--------------------------+-------------------------------+--------------------
+------------------------+-------------------------------+-----------------------------------
+------------------+----------------------+----------------------+--------------------------
+------------------+----------------------+--------+--------------------
+--------+--------------------+--------+---------------+--------+---------------
+------------------+----------------------+------------+--------------------
+------------+--------------------+------------+--------------------+------------
+--------------------+------------+--------------------+
|                   ID|LAST_UPDATED_DATE|LAST_UPDATED_TIMESTAMP|DUPLICATES|  POSTED|  EXPIRED
+--------------------+-----------------+----------------------+----------
+---------+---------+---------+--------------------+--------------------+--------------------
+-----------+--------------------+--------------------+--------------------
```

```
+--------------+---------------+---------+--------------------+--------------------
+------------------+--------------------+--------------------+-------------
+------------------+------------+-----------------+--------------+--------------------
+--------------------+--------------------+-------------+------+-----------
+----------------+------------------+---------+-----------+--------------------
+--------------------+------------------+------+--------------------+-----
+--------------------+-----+-------------+---------------+--------------------
+--------------+--------------------+-----------+--------------------
+-----------+--------------------+------+--------------------+------+--------------------
+------+--------------------+------+--------------------+------+--------------------
+------------------+--------------------+--------------------+--------------------
+--------------------+--------------------+----------------------+--------------------
+--------------------+--------------------+--------------------+--------------------
+--------------------+----------+--------------------+----------+--------------------
+--------------------+--------------------+--------------------+--------------------
+--------------------+--------------------+----------+--------------------
+----------+--------------------+----------+---------------+----------+---------------
+--------------+--------------------+--------------+--------------------
+--------------------------+--------------------------------+--------------------
+--------------------------+--------------------------------+----------------------------------
+------------------+----------------------+----------------------+--------------------------
+------------------+----------------------+-------+--------------------
+-------+--------------------+-------+---------------+-------+---------------
+------------------+----------------------+------------+--------------------
+-----------+--------------------+-----------+--------------------+------------
+--------------------+------------+--------------------+
|1f57d95acf4dc67ed...|         9/6/2024|  2024-09-06 20:32:...|         0|6/2/2024| 6/8/2024
May-2024\n\nEn...|       6/8/2024|               6|   894731|          Murphy USA|          N
time (> 32 h...|                  2|                  2|        false|  NULL|          0|
2051.01|Business Intellig...|15-2051.01|Business Intellig...|[\n  "45.0601",\n...|[\n  "Econ
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231010|Business Intellig..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|        [\n  7\n]|  [\n  "Artificial ...|          44|        Retail Tra
|0cb072af26757b6c4...|         8/2/2024|  2024-08-02 17:08:...|         0|6/2/2024| 8/1/2024
time (> 32 h...|                  3|                  3|        false|  NULL|          1|
Watervill...|   23|        Maine|          23011|        Kennebec, ME|          23011|
Watervill...|       12300|Augusta-Watervill...|    56|Administrative an...|   561|Administrat
2051.01|Business Intellig...|15-2051.01|Business Intellig...|                  []|
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231010|Business Intellig..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|             NULL|                  NULL|          56|Administrative an
```

```
|85318b12b3331fa49...|         9/6/2024|  2024-09-06 20:32:...|         1|6/2/2024| 7/7/2024
time (> 32 h...|                    5|                NULL|        false|  NULL|          0|
Fort Worth...|   48|        Texas|          48113|          Dallas, TX|          48113|
Fort Worth...|       19100|Dallas-Fort Worth...|    52|Finance and Insur...|   524|Insurance
2051.01|Business Intellig...|15-2051.01|Business Intellig...|                  []|
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231113|Data / Data Minin..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|             NULL|                  NULL|          52|Finance and Insur
|1b5c3941e54a1889e...|         9/6/2024|  2024-09-06 20:32:...|         1|6/2/2024|7/20/2024
time (> 32 h...|                    3|                NULL|        false|  NULL|          0|
Mesa-Chan...|    4|      Arizona|           4013|        Maricopa, AZ|           4013|
Mesa-Chan...|       38060|Phoenix-Mesa-Chan...|    52|Finance and Insur...|   522|Credit Int
2051.01|Business Intellig...|15-2051.01|Business Intellig...|                  []|
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231113|Data / Data Minin..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|        [\n  6\n]|  [\n  "Data Privac...|          52|Finance and Insur
|cb5ca25f02bdf25c1...|        6/19/2024|   2024-06-19 07:00:00|         0|6/2/2024|6/17/2024
time / full-...|                NULL|                NULL|        false| 92500|          0|
2051.01|Business Intellig...|15-2051.01|Business Intellig...|                  []|
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231010|Business Intellig..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|             NULL|                  NULL|          99|Unclassified Indu
|35a6cd2183d9fb270...|         9/6/2024|  2024-09-06 20:32:...|         0|6/2/2024|6/12/2024
time (> 32 h...|                NULL|                NULL|        false|110155|          1|
2051.01|Business Intellig...|15-2051.01|Business Intellig...|   [\n  "52.0201"\n]|[\n  "Busi
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231113|Data / Data Minin..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|             NULL|                  NULL|          51|         Informat
|06de8d192f30b1d8d...|         8/2/2024|  2024-08-02 17:08:...|         0|6/2/2024| 8/1/2024
time (> 32 h...|                NULL|                NULL|        false|  NULL|          0|
Mesa-Chan...|    4|      Arizona|           4013|        Maricopa, AZ|           4013|
Mesa-Chan...|       38060|Phoenix-Mesa-Chan...|    31|       Manufacturing|   334|Computer a
2051.01|Business Intellig...|15-2051.01|Business Intellig...|                  []|
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231113|Data / Data Minin..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|             NULL|                  NULL|          31|       Manufactur
|3d589c9d84677ca94...|         9/6/2024|  2024-09-06 20:32:...|         1|6/2/2024| 7/7/2024
time (> 32 h...|                    5|                NULL|        false|  NULL|          0|
```

```
Kettering, OH|   39|         Ohio|          39113|      Montgomery, OH|          39113|
Kettering, OH|       19430|Dayton-Kettering-...|    52|Finance and Insur...|   524|Insurance
2051.01|Business Intellig...|15-2051.01|Business Intellig...|                 []|
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231113|Data / Data Minin..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|             NULL|                  NULL|          52|Finance and Insur
|5a843df632e1ff756...|        6/21/2024|   2024-06-21 07:00:00|         0|6/2/2024|6/20/2024
time (> 32 h...|                   7|                   7|        false|  NULL|          0|
Newark-J...|   34|   New Jersey|          34037|          Sussex, NJ|          34037|
Newark-J...|       35620|New York-Newark-J...|    99|Unclassified Indu...|   999|Unclassifie
2051.01|Business Intellig...|15-2051.01|Business Intellig...|                 []|
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231010|Business Intellig..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|             NULL|                  NULL|          99|Unclassified Indu
|229620073766234e8...|        10/9/2024|  2024-10-09 18:07:...|         0|6/2/2024| 8/1/2024
time (> 32 h...|                   2|                   2|        false| 92962|          0|
Newark-J...|   36|     New York|          36061|        New York, NY|          36061|
Newark-J...|       35620|New York-Newark-J...|    54|Professional, Sci...|   541|Professional
2051.01|Business Intellig...|15-2051.01|Business Intellig...|[\n  "52.0101",\n...|[\n  "Busi
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231113|Data / Data Minin..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|        [\n  7\n]|  [\n  "Artificial ...|          54|Professional, Sci
|b7aa80a24c82f080c...|        9/28/2024|  2024-09-28 14:06:...|         8|6/2/2024|9/27/2024
time (> 32 h...|                  10|                NULL|        false|107645|          2|
Delan...|    42|     Wholesale Trade|   423|Merchant Wholesal...|  4238|Machinery, Equipm...
2051.01|Business Intellig...|15-2051.01|Business Intellig...|[\n  "14.0101",\n...|[\n  "Engi
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231113|Data / Data Minin..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|             NULL|                  NULL|          42|     Wholesale Tra
|2a107fd40bb1afac4...|        6/17/2024|   2024-06-17 07:00:00|         0|6/2/2024| 6/8/2024
time (> 32 h...|                   2|                NULL|        false|  NULL|          0|
2051.01|Business Intellig...|15-2051.01|Business Intellig...|[\n  "11.0701",\n...|[\n  "Compu
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231113|Data / Data Minin..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|             NULL|                  NULL|          56|Administrative an
|fd48c3ce533c3d20a...|         9/6/2024|  2024-09-06 20:32:...|         0|6/2/2024| 7/5/2024
time (> 32 h...|                NULL|                NULL|        false|  NULL|          0|
2051.01|Business Intellig...|15-2051.01|Business Intellig...|                 []|
```

```
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231113|Data / Data Minin..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|             NULL|                  NULL|          52|Finance and Insur
|57b527ea0f91db5bb...|         9/6/2024|  2024-09-06 20:32:...|         0|6/2/2024|7/27/2024
time (> 32 h...|                   6|                NULL|        false|192800|          0|
Warren-De...|   26|     Michigan|          26163|           Wayne, MI|          26163|
Warren-De...|       19820|Detroit-Warren-De...|    54|Professional, Sci...|   541|Professiona
2051.01|Business Intellig...|15-2051.01|Business Intellig...|   [\n  "45.0702"\n]|[\n  "Geogr
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231010|Business Intellig..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|        [\n  3\n]|  [\n  "Green Jobs:...|          54|Professional, Sci
|036cd733481fbcc98...|         8/2/2024|  2024-08-02 17:08:...|         0|6/2/2024| 8/1/2024
time (> 32 h...|                NULL|                NULL|        false| 81286|          1|
2051.01|Business Intellig...|15-2051.01|Business Intellig...|                  []|
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231510|Computer Systems ..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|             NULL|                  NULL|          52|Finance and Insur
|138ce2c9453b47a9b...|        8/10/2024|  2024-08-10 19:36:...|         5|6/2/2024| 8/9/2024
time (> 32 h...|                   5|                   5|        false|  NULL|          1|
Cambridge-...|   25|Massachusetts|          25025|         Suffolk, MA|          25025|
Cambridge-...|       14460|Boston-Cambridge-...|    61|Educational Services|   611|Educationa
2051.01|Business Intellig...|15-2051.01|Business Intellig...|   [\n  "52.0201"\n]|[\n  "Busin
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231113|Data / Data Minin..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|             NULL|                  NULL|          61|Educational Servic
|dd191e2ce3062c371...|         9/6/2024|  2024-09-06 20:32:...|         0|6/2/2024|6/20/2024
time (> 32 h...|                  12|                NULL|        false|125900|          0|
2051.01|Business Intellig...|15-2051.01|Business Intellig...|                  []|
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231010|Business Intellig..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|             NULL|                  NULL|          54|Professional, Sci
|99856b5a8a1c75d90...|         9/6/2024|  2024-09-06 20:32:...|         0|6/2/2024| 8/1/2024
time (> 32 h...|                   3|                   3|        false|  NULL|          1|
2051.01|Business Intellig...|15-2051.01|Business Intellig...|                  []|
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231010|Business Intellig..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|             NULL|                  NULL|          56|Administrative an
```

```
|f28123528a32b8c9b...|         9/6/2024|  2024-09-06 20:32:...|         0|6/2/2024| 8/1/2024
time (> 32 h...|                   8|                   8|        false|  NULL|          0|
Hoover...|    1|      Alabama|           1073|       Jefferson, AL|           1073|       Je
Hoover...|       13820|      Birmingham, AL|    62|Health Care and S...|   621|Ambulatory Hea
2051.01|Business Intellig...|15-2051.01|Business Intellig...|                  []|
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231510|Computer Systems ..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|        [\n  5\n]|  [\n  "Cybersecuri...|          62|Health Care and S
|b4e618e8d2a2b6744...|        10/9/2024|  2024-10-09 18:07:...|         2|6/2/2024|8/11/2024
time (> 32 h...|                   6|                NULL|        false|170000|          0|
Oak...|    6|   California|           6075|   San Francisco, CA|           6075|   San Franci
Oak...|       41860|San Francisco-Oak...|    99|Unclassified Indu...|   999|Unclassified Indu
2051.01|Business Intellig...|15-2051.01|Business Intellig...|                  []|
0000|Computer and Math...|   15-2000|Mathematical Scie...|   15-2050|Data Scientists|   15-
2051|Data Scientists|             23|Information Techn...|        231113|Data / Data Minin..
0000|Computer and Math...|15-2000|Mathematical Scie...|15-2050|Data Scientists|15-
2051|Data Scientists|        [\n  6\n]|  [\n  "Data Privac...|          99|Unclassified Indu
+--------------------+-----------------+----------------------+----------
+--------+---------+--------+--------------------+--------------------+--------------------
+----------+------------------+--------------------+--------------------
+--------------+----------------+---------+--------------------+--------------------
+------------------+--------------------+----------------------+-------------
+------------------+-------------+------------------+--------------+--------------------
+--------------------+--------------------+-------------+------+-----------
+----------------+------------------+---------+-----------+--------------------
+--------------------+------------------+------+--------------------+-----
+--------------------+-----+-------------+---------------+--------------------
+---------------+--------------------+------------+--------------------
+------------+--------------------+------+--------------------+------+--------------------
+------+--------------------+------+--------------------+------+--------------------
+------------------+--------------------+--------------------+--------------------
+--------------------+--------------------+----------------------+--------------------
+--------------------+--------------------+--------------------+--------------------
+--------------------+---------+--------------------+---------+--------------------
+--------------------+--------------------+--------------------+--------------------
+--------------------+--------------------+---------+--------------------
+---------+--------------------+---------+---------------+---------+---------------
+--------------+------------------+-------------+--------------------
+--------------------------+--------------------------------+--------------------
+--------------------------+--------------------------------+----------------------------------
+------------------+----------------------+----------------------+--------------------------
+------------------+----------------------+-------+--------------------
```

```
+-------+--------------------+-------+--------------+-------+--------------
+-----------------+----------------------+------------+--------------------
+------------+--------------------+------------+--------------------+------------
+--------------------+------------+--------------------+
only showing top 20 rows

root
 |-- SALARY: integer (nullable = true)
 |-- SALARY_FROM: integer (nullable = true)
 |-- SALARY_TO: integer (nullable = true)
 |-- MIN_YEARS_EXPERIENCE: integer (nullable = true)
 |-- MAX_YEARS_EXPERIENCE: integer (nullable = true)
 |-- DURATION: integer (nullable = true)
 |-- MODELED_DURATION: integer (nullable = true)
 |-- POSTED: string (nullable = true)
 |-- EXPIRED: string (nullable = true)
 |-- LAST_UPDATED_DATE: string (nullable = true)
 |-- MODELED_EXPIRED: date (nullable = true)
```

2 Part 2: Data Cleaning and Typecasting

3 Part 3: Salary by Education Level

4 Part 4: Salary by Remote Work Type