

# 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
from pyspark.sql.functions import 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","true")

# Verifying Data: Display Schema (column names & data types)
# print("---This is Diagnostic check, No need to print it in the final doc---") # Comment line
# 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 line
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()

# Registering DataFrame as Temporary SQL table
df_cleaned.createOrReplaceTempView("job_postings")
```

[Stage 7:>  $(o + 1) / 1]$

ID	LAST_UPDATED_DATE	LAST_UPDATED_TIMESTAMP	DUPLICATES	POSTED	EXPIRED
1	2023-10-01 12:00:00	2023-10-01 12:00:00	0	True	False
2	2023-10-02 13:00:00	2023-10-02 13:00:00	1	True	False
3	2023-10-03 14:00:00	2023-10-03 14:00:00	0	True	False
4	2023-10-04 15:00:00	2023-10-04 15:00:00	1	True	False
5	2023-10-05 16:00:00	2023-10-05 16:00:00	0	True	False
6	2023-10-06 17:00:00	2023-10-06 17:00:00	1	True	False
7	2023-10-07 18:00:00	2023-10-07 18:00:00	0	True	False
8	2023-10-08 19:00:00	2023-10-08 19:00:00	1	True	False
9	2023-10-09 20:00:00	2023-10-09 20:00:00	0	True	False
10	2023-10-10 21:00:00	2023-10-10 21:00:00	1	True	False
11	2023-10-11 22:00:00	2023-10-11 22:00:00	0	True	False
12	2023-10-12 23:00:00	2023-10-12 23:00:00	1	True	False
13	2023-10-13 00:00:00	2023-10-13 00:00:00	0	True	False
14	2023-10-14 01:00:00	2023-10-14 01:00:00	1	True	False
15	2023-10-15 02:00:00	2023-10-15 02:00:00	0	True	False
16	2023-10-16 03:00:00	2023-10-16 03:00:00	1	True	False
17	2023-10-17 04:00:00	2023-10-17 04:00:00	0	True	False
18	2023-10-18 05:00:00	2023-10-18 05:00:00	1	True	False
19	2023-10-19 06:00:00	2023-10-19 06:00:00	0	True	False
20	2023-10-20 07:00:00	2023-10-20 07:00:00	1	True	False
21	2023-10-21 08:00:00	2023-10-21 08:00:00	0	True	False
22	2023-10-22 09:00:00	2023-10-22 09:00:00	1	True	False
23	2023-10-23 10:00:00	2023-10-23 10:00:00	0	True	False
24	2023-10-24 11:00:00	2023-10-24 11:00:00	1	True	False
25	2023-10-25 12:00:00	2023-10-25 12:00:00	0	True	False
26	2023-10-26 13:00:00	2023-10-26 13:00:00	1	True	False
27	2023-10-27 14:00:00	2023-10-27 14:00:00	0	True	False
28	2023-10-28 15:00:00	2023-10-28 15:00:00	1	True	False
29	2023-10-29 16:00:00	2023-10-29 16:00:00	0	True	False
30	2023-10-30 17:00:00	2023-10-30 17:00:00	1	True	False
31	2023-10-31 18:00:00	2023-10-31 18:00:00	0	True	False
32	2023-11-01 19:00:00	2023-11-01 19:00:00	1	True	False
33	2023-11-02 20:00:00	2023-11-02 20:00:00	0	True	False
34	2023-11-03 21:00:00	2023-11-03 21:00:00	1	True	False
35	2023-11-04 22:00:00	2023-11-04 22:00:00	0	True	False
36	2023-11-05 23:00:00	2023-11-05 23:00:00	1	True	False
37	2023-11-06 00:00:00	2023-11-06 00:00:00	0	True	False
38	2023-11-07 01:00:00	2023-11-07 01:00:00	1	True	False
39	2023-11-08 02:00:00	2023-11-08 02:00:00	0	True	False
40	2023-11-09 03:00:00	2023-11-09 03:00:00	1	True	False
41	2023-11-10 04:00:00	2023-11-10 04:00:00	0	True	False
42	2023-11-11 05:00:00	2023-11-11 05:00:00	1	True	False
43	2023-11-12 06:00:00	2023-11-12 06:00:00	0	True	False
44	2023-11-13 07:00:00	2023-11-13 07:00:00	1	True	False
45	2023-11-14 08:00:00	2023-11-14 08:00:00	0	True	False
46	2023-11-15 09:00:00	2023-11-15 09:00:00	1	True	False
47	2023-11-16 10:00:00	2023-11-16 10:00:00	0	True	False
48	2023-11-17 11:00:00	2023-11-17 11:00:00	1	True	False
49	2023-11-18 12:00:00	2023-11-18 12:00:00	0	True	False
50	2023-11-19 13:00:00	2023-11-19 13:00:00	1	True	False
51	2023-11-20 14:00:00	2023-11-20 14:00:00	0	True	False
52	2023-11-21 15:00:00	2023-11-21 15:00:00	1	True	False
53	2023-11-22 16:00:00	2023-11-22 16:00:00	0	True	False
54	2023-11-23 17:00:00	2023-11-23 17:00:00	1	True	False
55	2023-11-24 18:00:00	2023-11-24 18:00:00	0	True	False
56	2023-11-25 19:00:00	2023-11-25 19:00:00	1	True	False
57	2023-11-26 20:00:00	2023-11-26 20:00:00	0	True	False
58	2023-11-27 21:00:00	2023-11-27 21:00:00	1	True	False
59	2023-11-28 22:00:00	2023-11-28 22:00:00	0	True	False
60	2023-11-29 23:00:00	2023-11-29 23:00:00	1	True	False
61	2023-11-30 00:00:00	2023-11-30 00:00:00	0	True	False
62	2023-12-01 01:00:00	2023-12-01 01:00:00	1	True	False
63	2023-12-02 02:00:00	2023-12-02 02:00:00	0	True	False
64	2023-12-03 03:00:00	2023-12-03 03:00:00	1	True	False
65	2023-12-04 04:00:00	2023-12-04 04:00:00	0	True	False
66	2023-12-05 05:00:00	2023-12-05 05:00:00	1	True	False
67	2023-12-06 06:00:00	2023-12-06 06:00:00	0	True	False
68	2023-12-07 07:00:00	2023-12-07 07:00:00	1	True	False
69	2023-12-08 08:00:00	2023-12-08 08:00:00	0	True	False
70	2023-12-09 09:00:00	2023-12-09 09:00:00	1	True	False
71	2023-12-10 10:00:00	2023-12-10 10:00:00	0	True	False
72	2023-12-11 11:00:00	2023-12-11 11:00:00	1	True	False
73	2023-12-12 12:00:00	2023-12-12 12:00:00	0	True	False
74	2023-12-13 13:00:00	2023-12-13 13:00:00	1	True	False
75	2023-12-14 14:00:00	2023-12-14 14:00:00	0	True	False
76	2023-12-15 15:00:00	2023-12-15 15:00:00	1	True	False
77	2023-12-16 16:00:00	2023-12-16 16:00:00	0	True	False
78	2023-12-17 17:00:00	2023-12-17 17:00:00	1	True	False
79	2023-12-18 18:00:00	2023-12-18 18:00:00	0	True	False
80	2023-12-19 19:00:00	2023-12-19 19:00:00	1	True	False
81	2023-12-20 20:00:00	2023-12-20 20:00:00	0	True	False
82	2023-12-21 21:00:00	2023-12-21 21:00:00	1	True	False
83	2023-12-22 22:00:00	2023-12-22 22:00:00	0	True	False
84	2023-12-23 23:00:00	2023-12-23 23:00:00	1	True	False
85	2023-12-24 00:00:00	2023-12-24 00:00:00	0	True	False
86	2023-12-25 01:00:00	2023-12-25 01:00:00	1	True	False
87	2023-12-26 02:00:00	2023-12-26 02:00:00	0	True	False
88	2023-12-27 03:00:00	2023-12-27 03:00:00	1	True	False
89	2023-12-28 04:00:00	2023-12-28 04:00:00	0	True	False
90	2023-12-29 05:00:00	2023-12-29 05:00:00	1	True	False
91	2023-12-30 06:00:00	2023-12-30 06:00:00	0	True	False
92	2023-12-31 07:00:00	2023-12-31 07:00:00	1	True	False
93	2024-01-01 08:00:00	2024-01-01 08:00:00	0	True	False
94	2024-01-02 09:00:00	2024-01-02 09:00:00	1	True	False
95	2024-01-03 10:00:00	2024-01-03 10:00:00	0	True	False
96	2024-01-04 11:00:00	2024-01-04 11:00:00	1	True	False
97	2024-01-05 12:00:00	2024-01-05 12:00:00	0	True	False
98	2024-01-06 13:00:00	2024-01-06 13:00:00	1	True	False
99	2024-01-07 14:00:00	2024-01-07 14:00:00	0	True	False
100	2024-01-08 15:00:00	2024-01-08 15:00:00	1	True	False
101	2024-01-09 16:00:00	2024-01-09 16:00:00	0	True	False
102	2024-01-10 17:00:00	2024-01-10 17:00:00	1	True	False
103	2024-01-11 18:00:00	2024-01-11 18:00:00	0	True	False
104	2024-01-12 19:00:00	2024-01-12 19:00:00	1	True	False
105	2024-01-13 20:00:00	2024-01-13 20:00:00	0	True	False
106	2024-01-14 21:00:00	2024-01-14 21:00:00	1	True	False
107	2024-01-15 22:00:00	2024-01-15 22:00:00	0	True	False
108	2024-01-16 23:00:00	2024-01-16 23:00:00	1	True	False
109	2024-01-17 00:00:00	2024-01-17 00:00:00	0	True	False
110	2024-01-18 01:00:00	2024-01-18 01:00:00	1	True	False
111	2024-01-19 02:00:00	2024-01-19 02:00:00	0	True	False
112	2024-01-20 03:00:00	2024-01-20 03:00:00	1	True	False
113	2024-01-21 04:00:00	2024-01-21 04:00:00	0	True	False
114	2024-01-22 05:00:00	2024-01-22 05:00:00	1	True	False
115	2024-01-23 06:00:00	2024-01-23 06:00:00	0	True	False
116	2024-01-24 07:00:00	2024-01-24 07:00:00	1	True	False
117	2024-01-25 08:00:00	2024-01-25 08:00:00	0	True	False
118	2024-01-26 09:00:00	2024-01-26 09:00:00	1	True	False
119	2024-01-27 10:00:00	2024-01-27 10:00:00	0	True	False
120	2024-01-28 11:00:00	2024-01-28 11:00:00	1	True	False
121	2024-01-29 12:00:00	2024-01-29 12:00:00	0	True	False
122	2024-01-30 13:00:00	2024-01-30 13:00:00	1	True	False
123	2024-01-31 14:00:00	2024-01-31 14:00:00	0	True	False
124	2024-02-01 15:00:00	2024-02-01 15:00:00	1	True	False
125	2024-02-02 16:00:00	2024-02-02 16:00:00	0	True	False
126	2024-02-03 17:00:00	2024-02-03 17:00:00	1	True	False
127	2024-02-04 18:00:00	2024-02-04 18:00:00	0	True	False
128	2024-02-05 19:00:00	2024-02-05 19:00:00	1	True	False
129	2024-02-06 20:00:00	2024-02-06 20:00:00	0	True	False
130	2024-02-07 21:00:00	2024-02-07 21:00:00	1	True	False
131	2024-02-08 22:00:00	2024-02-08 22:00:00	0	True	False
132	2024-02-09 23:00:00	2024-02-09 23:00:00	1	True	False
133	2024-02-10 00:00:00	2024-02-10 00:00:00	0	True	False
134	2024-02-11 01:00:00	2024-02-11 01:00:00	1	True	False
135	2024-02-12 02:00:00	2024-02-12 02:00:00	0	True	False
136	2024-02-13 03:00:00	2024-02-13 03:00:00	1	True	False
137	2024-02-14 04:00:00	2024-02-14 04:00:00	0	True	False
138	2024-02-15 05:00:00	2024-02-15 05:00:00	1	True	False
139	2024-02-16 06:00:00	2024-02-16 06:00:00	0	True	False
140	2024-02-17 07:00:00	2024-02-17 07:00:00	1	True	False
141	2024-02-18 08:00:00	2024-02-18 08:00:00	0	True	False
142	2024-02-19 09:00:00	2024-02-19 09:00:00	1	True	False
143	2024-02-20 10:00:00	2024-02-20 10:00:00	0	True	False
144	2024-02-21 11:00:00</				

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	1
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	
Waterville...  23  Maine  23011  Kennebec, ME  23011					
Waterville...  12300 Augusta-Waterville...  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  Information
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 an			
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 Unclassified			
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 Professiona			

2051.01 Business Intellig...	15-2051.01 Business Intellig...	[\"52.0101\",\n...]	[\"Business 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... [\"14.0101\",\n...] [\n "Engin 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... [\"11.0701\",\n...] [\n "Comp 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 Profession 2051.01 Business Intellig... 15-2051.01 Business Intellig...  [\n "45.0702"\n] [\n "Geog 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
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 "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	61 Educational Servi	
dd191e2ce3062c371...	9/6/2024	2024-09-06 20:32:...	0 6/2/2024	6/20/2024
time (> 32 h...	12	NULL	false	125900
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
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
Hoover...	1  Alabama	1073	Jefferson, AL	1073
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
Oak...	6  California	6075	San Francisco, CA	6075
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		

```
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**