

REMINDER: The main focus of the exercise is on working with messy data

QUESTIONS

Who gets hired?

What kind of **talent** do **employers want** when they are hiring a data scientist?

Which **location** has the most opportunities?

What **skills, tools, degrees or majors** do **employers want** the most for data scientists?

WHAT TO INCLUDE

Insights, commentary

- insight to include the **ability to summarise results** by **job profile**, -
company, location city, area of the country

Review of **how easy the data was to work with**
highlighting **any limitations** you found in the data set.

The **insights** you find can be **documented simply** with:

- screenshots** of your SQL query results/previews in your SQL client (or if you feel confident using them, from data frames and matplotlib /seaborn (saved out as images) from your python jupyter notebook)
- some **simple annotation/text summarising**:
 - what you found
 - how easy it was to get there

Employ **string functions** + regexp, Like , case statements to extract common values from strings of different lengths, eg job description

Create **new columns**:

- employ **Boolean T/F logic** to indicate which roles are **closest to big financial or software centres** in the US

Maybe a case Statement. Define big financial, software
Centres

Seattle, San Francisco, Los Angeles, NEw York, Chicago,

Make a **decision** about **handling NULLs** in the data - fill in values where logical, ignore them or clean them where not

As you have **ONLY one page** to make your case, you might start by:

- identifying multiple trends**

- then scale back to focus on just one or two.,

- If you don't find any great data insights, you should feel free to take **screen shots of your cleaning procedures** and talk about them.

EXPECTED OUTCOME:

1. **one data source** which combines all the input csvs that is usable for rapidly running queries to find insights.

The data can be in **any format** you have been working with in the course so far - **pandas data frame, csv flat file, MySQL table or MySQL database**

2. **'one pager' summary**

This can be in **pdf, slide, word doc etc...** this can be as beautiful or as simple as you like. You will be sharing this with your classmates and the teaching team will provide feedback on your submissions next week.