End to End Use Case

Resources: https://github.com/fenago/cts245X/tree/main/jobs

Lab 1:

Connect to the `job_postings.csv` table from the `Datasets` folder.

Rename the table to `Job Postings`.

Go to the Power Query Editor and investigate the missing values for `Company Name` by hovering over the header title.

What is the percent of missing values for the Minimum and Maximum pay columns?

Lab 2:

The Power Query Editor is very powerful for performing initial exploratory data analysis (EDA). Let's use it's built-in functionality to explore basic statistics of our dataset.

We do not recommend doing so, but if you lost progress you can load the solution workbook of the previous exercise `1_1_import_data_solution.pbix` from the Exercises folder.

Change column profiling from the top 1000 rows to be based on the entire dataset. You can do so in the left bottom corner.

Let's investigate columns with unique values. Under the *View* ribbon, select *Column distribution* to display distinct and unique values for each column.

Let's dive even deeper into the columns by looking at *Column statistics* and *Value distribution*, by navigating to the *View* ribbon to enable *Column profile*.

Before proceeding, take a few minutes to explore the columns more by clicking the associated column header.

How many distinct values can be found in the company industry column?

What is the second most popular industry?

Lab 3:

Let's explore trends in job postings over time; specifically, we want to look at how the years of experience relate to a particular job position (e.g., entry-level).

We do not recommend doing so, but if you lost progress you can load the solution workbook of the previous exercise `1_2_initial_eda_solution.pbix` from the Exercises folder.

Close the Power Query Editor, in case it is still open, and rename "Page 1" to "Job Level Analysis".

To have an initial idea of the relation between time and position level, visualize the average of `Years of Experience` required by each of the fields of `Job Position Level`.

Now, to have a better insight let's visualize how the number of job postings change over time using the `Job Posting Date` column. Use job position level to differentiate these postings visually.

What month and year has the least amount of job postings? (e.g., October 2016)

Lab 4:

It's time to drill down into the 5 jobs of interest: Data Scientist, Data Analyst, Data Engineer, Machine Learning Engineer, and Data Science Manager.

DataSearch has requested specific analysis of these top jobs in data science. So let's clean up some data!

We do not recommend doing so, but if you lost progress you can load the solution workbook of the previous exercise `1_3_posting_trends_solution.pbix` from the Exercises folder.

Create a new "Job Titles" page and visualize the counts of all the different job titles compared to each other.

For this we recommend using a *Treemap*. Later in the case study, it will be your job to find the fitting visualization for a problem.

Filter down **all pages** to focus on the relevant job title fields of: "Data Analyst", "Data Engineer", "Data Scientist", "Machine Learning Engineer", "Data Science Manager".

How many job postings are listed for the second highest number of jobs for the filtered values?

Lab 5:

Let's investigate what is influencing job market trends. For this, we will be looking at different job's salaries and analyzing how they change with years of experience.

Since the dataset only provides minimum and maximum values for salaries, we will need to do some data cleanup first.

We do not recommend doing so, but if you lost progress you can load the solution workbook of the previous exercise `1_4_data_cleanup_solution.pbix` from the Exercises folder.

Estimate the average salary by creating an `Average Pay` calculated column. Take a moment and think about how you can do this using `Minimum Pay` and `Maximum Pay`.

Format `Average Pay` as dollars with zero decimal places.

We are about to start creating measures with DAX and need an organized way to store them. Create a new table titled `_Measures`

Create a measure for the average of `Average Pay` in the `_Measures` table, and name it `Average of Average Pay`. Do not forget to format it in dollars.

Create a new "Salary Analysis" page and visualize this new measure of `Average of Average Pay` versus years of experience for each job title.

Let's understand the overall trends of pay versus experience better by adding a trend lines for each `Job Title`.

Which `Job Title` has the lowest average pay based on experience level?

- ○ Machine Learning Engineer
- ○ Data Scientist
- ○ Data Engineer
- ○ Data Analyst

Lab 6:

Let's jump into analyzing skills. The `Job Skills` column contains values in a list (e.g., ['python', 'power_bi', ...]), so we'll need to clean this up before further analysis.

We do not recommend doing so, but if you lost progress you can load the solution workbook of the previous exercise `1_5_experience_vs_salary_solution.pbix` from the Exercises folder.

With the Power Query Editor, create a new query with a duplicate of our current `Job Posting` table.

Label this new table `Job Skills` and remove all columns except `Job Posting ID` & `Job Skills`.

Clean up the `Job Skills` column to remove brackets (e.g., []), quotes (e.g, ''), and spaces (e.g., ' '). Only the comma should remains as the delimiter.

Split the `Job Skills` column by the delimiter to make a row for each skill along with the associated `Job Posting ID`.

Remove rows that have a blank value in the `Job Skills` column by using the down arrow in the header of this column.

Close & Apply this query. On a new page titled "Skill Analysis", create a visualization showing the counts of all the different skills.

How many times does Power BI appear as a skill in job postings?

- 419
- 234
- 653
- 1298

Lab 7:

Let's use the power of DAX to better understand what is the percent chance that a certain skill will be listed in a job posting.

We do not recommend doing so, but if you lost progress you can load the solution workbook of the previous exercise `2_1_skill_cleanup_solution.pbix` from the Exercises folder.

In the Power Query Editor, for the `Job Skills` column, ensure "power_bi" and "powerbi" are categorized under the same skill of "power_bi".

Exit the Power Query Editor. Create a measure, called `Skill Count`, that counts the skills in the `Job Skills` table.

On a new page, called "Skill Likelihood", create a matrix to visualize this newly created measure.

Create a measure, called `Posting Count` , for the count of all the job postings.

Make sure that your measure is added to the matrix and does the calculation properly.

Create our final measure, called `% Skill in Posting` , where we see the likelihood that a skill is within a posting.

Format the results as a percentage and verify with the matrix.

Take the matrix one step further by adding `Job Title` .

What percent likelihood is Power BI listed as a skill in a Data Scientist job posting?

Lab 8:

Let's put our new measure to good use by exploring how skills in job postings trend over time.

We do not recommend doing so, but if you lost progress you can load the solution workbook of the previous exercise `2_2_skill_likelihood_solution.pbix` from the Exercises folder.

Before we begin, cleanup the matrix from the last exercise. Remove `Skill Count` , `Posting Count` , and *Row subtotals*.

Create a time-series graph that shows the `% Skill in Posting` over time, with it displaying each skill for the legend.

This visualization is cluttered. Filter this graph to only show the top 10 results for `% Skill in Posting` . Add filters and slicers as necessary for the next question.

For the three major jobs (i.e., Data Analyst, Engineer, & Scientist), how are the top 10 skills trending over time?

- ○ Trending upwards
- ○ Maintaining at consistent demand
- ○ Trending downwards

Lab 9:

For *DataSearch*, we need to explore better the top companies and industries looking for data science roles. So let's dive in!

We do not recommend doing so, but if you lost progress you can load the solution workbook of the previous exercise `2_3_time_skills_solution.pbix` *from the Exercises folder.*

On a new page, titled "Experience Analysis", take some time to further explore the data by creating visualizations using different combinations of the following:

`'Job Postings'[Years of Experience]`
`'Job Postings'[Company Industry]`
`'Job Postings'[Company Size]`
`'Job Skills'[Job Skills]`

If not done in the previous step, create a visualization that shows the average years of experience by company industry, using a *Scatter chart*.

Fine tune this further, by showcasing both `Posting Count` and `Job Position Level` in this visualization.

Filter down to the top 10 industries by number of job postings.

Create a table, to supplement this visual. For this, show for different job titles, the companies and amount of job postings they have.

The "Internet" industry is hiring for many "Mid-Senior level" positions. What is the main company hiring for these positions?

Lab 10:

Alright, so we've done a very thorough analysis for our 5 key job titles that *DataSearch* suggested. But are there other job titles that we also should consider that have similar experience requirements?

We do not recommend doing so, but if you lost progress you can load the solution workbook of the previous exercise `2_4_description_details_solution.pbix` *from the Exercises folder.*

Go to the "Job Titles" page and update this *Treemap* visualization so that it shows counts of different skill requirements for each of the job titles.

It looks like not all skills are appearing on this *Treemap*. Filter this visualization in order to show the top 20 skills by `Skill Count`.

Remove the filters on all pages so we can investigate other job titles. Add some slicers for job titles and skills.

Take some time to find similar jobs, in regards to skills, for our previous job titles:

- Data Engineer
- Data Analyst
- Data Scientist
- Machine Learning Engineer
- Data Science Manager

What job is most similar to a Data Analyst in regards to the type and count of skills required?

- ○ Software Engineer
- ○ Developer
- ○ Software Development Engineer
- ○ Business Analyst