<u>**Summary Section**</u>

<u>**What actions should an inexperienced inspiring analyst take?**</u>

What skills to learn first? (List top 3 skills from most popular to least popular)

Database: Microsoft SQL Server, MySQL, PostgreSQL

Platform: AWS, Microsoft Azure, Google Cloud

Programming Language: SQL, Python, JavaScript

Which country has the highest employment rate for inexperienced analyst with less than 2 years of coding experience? Relocation should be considered if you aren't getting any luck in your current location. (List from highest to lowest)

Brazil, India, United Kingdom, United States of America, Canada, Germany, Australia.

Does education matters?

Not so much, you should focus on building your skills and experience because the results showed most employed analyst had lower education.

Which working model has the highest employment rate with less than 2 years of coding experience? (List from highest to lowest)

In-person, Remote, Hybrid.

This indicates that most companies still prefer in-person applicants for inexperienced analyst than Remote and Hybrid. You should apply for all the in-person jobs first, then apply for remote and hybrid positions to increase your chance of getting hired.

What is the employment success rate in percentage worldwide with less than 2 years of coding experience?

2.10%, which is tough for inexperienced analyst. Make sure you know what you are getting yourself into and have a good mindset coming into it.

Source from: https://insights.stackoverflow.com/survey

The survey was fielded from May 11, 2022 to June 1, 2022

This data contains 5635211 counts of data. (Over 5 million)

Click this link for the raw source and the cleaned dataset:

Data Cleaning

Excel:

1. Replaced 54080 corrupted punctuation (â€) with the correct punctuation (').

   Ex: Masterâ€™ s → Master's.

2. Exported as a workbook and imported into Microsoft SQL Server.

Microsoft SQL Server:

1. Removed duplicates using row_number/window function.

2. Replaced NULL value with 0 for YearsCode column because there should be 0 to indicate a candidate without any experience in coding, but it only has NA, 1, 2, etc. Therefore, I assume NA means 0 years of experience, this will make SQL coding and visualization in Tableau easier.

3. I couldn't populate any columns with NULL values because there wasn't a reference point.

4. Created two new tables: a Company_Table and a Background_Table, from the cleaned dataset. Response ID is kept for each table, so join function could be used after.

   - Company_Table has columns such as Response ID as the primary key, organization size, RemoteWork (working model), total compensation, compensation frequency.

   - Background_Table has columns such as Response ID as the primary key, education level, country, DevType (current job name), employment status, database have worked with, age, gender, sex, etc.

Data Exploration

Microsoft SQL Server:

You can find the actual SQL exploration queries with detailed explanations of the edge cases that I also covered from this link:
https://github.com/HaomingChen1998/Portfolio-Project/blob/main/Analyst%20Employment%20Rate/Data_Analyst_Project.sql

Things I discovered were:

1. Most popular/common Platform, Language, and Database skill usage counts for an employed person as a data analyst worldwide.

2. Analyst employment rate per country with less than 2 year of coding experience (any coding education also counts toward coding experience).

3. Analyst employment rate per education level.

4. Analyst employment rate per working model (remote/hybrid/in-person) with less than 2 year of coding experience.

5. Analyst global employment rate with less than 2 year of coding experience.

Tableau:

Visualization can be found at this link:
https://public.tableau.com/app/profile/haoming.chen1867/viz/FactorsofDataAnalystApplicantsEmploymentRate_/Dashboard1?publish=yes

What I Have Learned:

Advice from a <u>Senior Data Analytics Manager at Wunderkind</u> on my project:

1. Perform data cleaning such as removing duplicates and populate null values first then break up the data set.

2. Minimize the use of CTE or subquery.

Lesson learned from doing this project from my own mistakes and advice from Senior Data Analyst on LinkedIn.

1. Remove the duplicates and populate any null values if you have a reference point BEFORE you break up your dataset into different columns/ tables.

2. Before you remove the duplicate, use select statement to see what you're deleting, once you confirmed that everything looks correct, then change it to delete statement.

3. Do not start query before you fully understand the dataset you are working with.

4. Although I can remove duplicates much easier in Excel, but when the dataset becomes too large, Excel might crash while removing the duplicates, so SQL is better in terms of handling larger dataset.

5. When I want to see the employment rate per country, I need to exclude countries where the sample size is small. Because the results won't be significant, and they can skew the overall trend.

6. Before importing to Tableau, make sure to replace all the NULL values with 0 because it could mistake that with a string.