Daily Diary

2024-06-20

Pandas E-Commerce Worksheet

1. Imported the pandas library.
2. Read the "Ecommerce Purchases" CSV file and stored it in a dataframe.
3. Checked the first few rows of the dataframe to ensure it was loaded correctly.
4. Determined the shape of the dataframe, finding it has 10,000 rows and 14 columns.
5. Calculated the average purchase price, which is approximately $50.35.
6. Identified the highest purchase price as $99.99 and the lowest as $0.00.
7. Counted the number of people who chose English ('en') as their language, totaling 1,098.
8. Found out that 30 people have the job title of Lawyer.
9. Determined the number of purchases made during AM and PM, with 5,068 transactions.
10. Listed the five most common job titles:
    - Interior and spatial designer
    - Lawyer
    - Social researcher
    - Purchasing manager
    - Designer, jewellery
11. Discovered that the purchase price for the transaction from Lot "90 WT" was $75.1.
12. Found that 39 people used American Express for purchases above $95.
13. Determined that 1,033 people have credit cards expiring in 2025.
14. Identified the top five most popular email providers, with hotmail.com being the most common.

Pandas Salaries Worksheet

1. Imported the pandas library.
2. Read the "Salaries" CSV file and stored it in a dataframe.
3. Verified the dataframe was loaded correctly by checking the first few rows.
4. Listed the columns present in the dataframe, noting there are 13 columns.
5. Confirmed the dataframe contains 148,654 rows.
6. Displayed information about the dataframe, identifying columns with missing values.
7. Calculated the total BasePay, which amounts to approximately $9.82 billion.
8. Found the highest amount of overtime pay recorded.
9. Identified the job title of JOSEPH DRISCOLL as Captain, Fire Suppression.
10. Calculated JOSEPH DRISCOLL's total compensation, including benefits, to be $540,649.82.

11. Discovered that NATHANIEL FORD is the highest-paid person, including benefits.
12. Determined the average BasePay for all employees from 2011 to 2014.
13. Replaced missing values in the Benefits column with 0.
14. Identified there are 1,037 unique job titles in the dataframe.