Ace Digital, Ace Resources Advisory Sdn Bhd

Robotics Process Automation Pre Screen Technical Assessment

1 Question

1.1 Problem Description

The major catalyst for Crude Palm Oil (CPO) price movement is Soy Bean Oil (SBO) price direction. An organization requires the pricing of Palm Oil commodity to be compared with Soy Bean Oil and plotted in a Microsoft Excel worksheet. CPO vs SBO pricing can be extracted from the url: http://mpoc.org.my/daily-palm-oil-price/. Figure 1 illustrates a screen shot of the table containing the CPO and SBO price in US$ per metric tonne.

Date	CPO, BMD (US$/MT)	SBO, ROTT (US$/MT)	PO Discount to SBO
10 Sep 20	675	880	205
9 Sep 20	674	877	203
8 Sep 20	689	890	201
7 Sep 20	676	886	210
4 Sep 20	683	888	205
3 Sep 20	698	880	182
2 Sep 20	677	870	193
1 Sep 20	674	881	207

Figure 1: A screenshot of the 2nd table that can be found in the url: http://mpoc.org.my/daily-palm-oil-price/. The values shown in the table may differ based on the date when the web site is visited.

1.2 Design requirement

A bot need to be designed to fulfill the above requirement on daily basis. Any commercial RPA tools such as UiPath, Automation Anywhere, Blue Prism or open source based RPA tools such as Robot Framework, RPA Framework or TagUI can be used to solve the problem. Important steps are detailed below:

- Scrap the content of table shown in Figure 1 and store into a file (txt, csv or xlsx).

- Extract the values in the columns 'CPO, BMD (US$/MT)' and 'SBO, ROT (US$/MT)' and calculate 'Price Premium SBO over CPO' in a worksheet of Ms Excel.
- Plot 'Price Premium SBO over CPO' against its respective dates, within the same worksheet.

2 Question

2.1 Problem Description

Feed the paragraph content stored in a text file named 'paragraph.txt', shown in Figure 2.

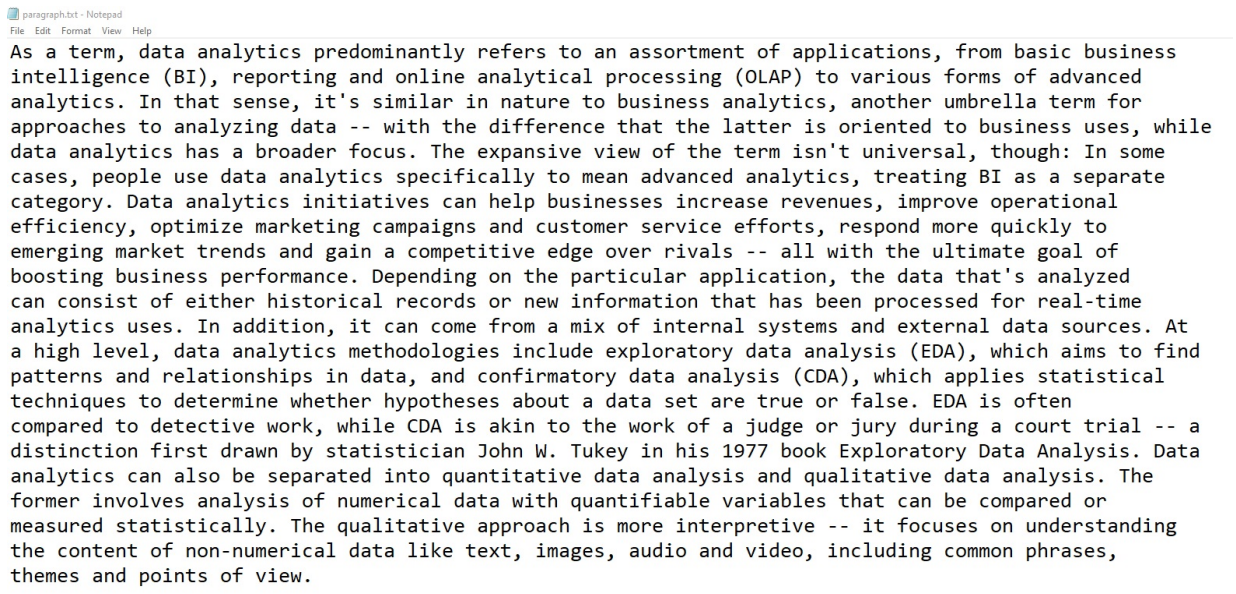
paragraph.txt - Notepad
File Edit Format View Help
As a term, data analytics predominantly refers to an assortment of applications, from basic business intelligence (BI), reporting and online analytical processing (OLAP) to various forms of advanced analytics. In that sense, it's similar in nature to business analytics, another umbrella term for approaches to analyzing data -- with the difference that the latter is oriented to business uses, while data analytics has a broader focus. The expansive view of the term isn't universal, though: In some cases, people use data analytics specifically to mean advanced analytics, treating BI as a separate category. Data analytics initiatives can help businesses increase revenues, improve operational efficiency, optimize marketing campaigns and customer service efforts, respond more quickly to emerging market trends and gain a competitive edge over rivals -- all with the ultimate goal of boosting business performance. Depending on the particular application, the data that's analyzed can consist of either historical records or new information that has been processed for real-time analytics uses. In addition, it can come from a mix of internal systems and external data sources. At a high level, data analytics methodologies include exploratory data analysis (EDA), which aims to find patterns and relationships in data, and confirmatory data analysis (CDA), which applies statistical techniques to determine whether hypotheses about a data set are true or false. EDA is often compared to detective work, while CDA is akin to the work of a judge or jury during a court trial -- a distinction first drawn by statistician John W. Tukey in his 1977 book Exploratory Data Analysis. Data analytics can also be separated into quantitative data analysis and qualitative data analysis. The former involves analysis of numerical data with quantifiable variables that can be compared or measured statistically. The qualitative approach is more interpretive -- it focuses on understanding the content of non-numerical data like text, images, audio and video, including common phrases, themes and points of view.

Figure 2: Content of the text file, 'paragraph.txt'

2.2 Design requirement

Solve the given questions using **Python** or **.NET** based data analytic packages.

- What is the probability of the word "data" occurring in each line (in total there are 22 lines) ?
- What is the distribution of distinct word counts across all the 22 lines ?
- What is the probability of the word "analytics" occurring after the word "data" ?

Note: Do not solve the above problem sentence by sentence.

3 Plagiarism Policy

Do not, under any circumstances, assign another person to work on the development tasks. **`Originality of the work will be verified during interview.`** You may use libraries, modules or functions published by other programmers in Github or other repositories.

4 Deliverable

- Capture screen shot of the process/program flow, outputs and other relevant evidences.
- Save those captured images in a Word Document.
- Explain briefly on each image. Quality of written communication is essential.

5 Duration

Complete the given task in 5 (FIVE) days.

6 Method of delivery

Items listed in Section 4 can be uploaded to a cloud based storage that can be easily downloaded or compress it into a single file and email the details to:

Mr Rafi Basha Shaik,
Asst. Manager - Talent Acquisition (Digital)
Center Support Human Resources
Email: Rafi_Shaik@aceras.biz
Tel: 03-27858888 (ext - 7707)
Mobile: +6011-33034625