**Pizza Sales Analysis**

Random CSV file: Data for admission in the university

Url: <https://www.kaggle.com/datasets/akshaydattatraykhare/data-for-admission-in-the-university>

**Introduction**

The pizza sales dataset consists of order id, order details id, pizza id, quantity, order date, order time, unit price, total price, pizza size, pizza type, pizza ingredients, and pizza name.

The data for admission in the university dataset consists of GRE score, TOEFL score, university rating, SOP (Statement of Purpose), LOR (Letter of Recommendation), CGPA, research, and the chance of admission.

The dataset used for Task 1 does not have the same attributes as the university dataset obtained randomly from Kaggle. In order to run the code developed for Task 1 on this dataset successfully, the entire column name and position must be changed.

Using the import csv module, the new csv file (admission data) is successfully imported into the Jupyter notebook. [Fig. 1]. When trying to execute task 1's code on the admission dataset, an error message stating that "total price" is not in the list is displayed. [Fig. 2]. Table 1 contrasts the time required to complete Tasks 1 and 2 with the size of the data set at its default size and after it has been intentionally increased.

**What I learnt about applying Python to data analysis**

Through this assignment, I discovered how simple it is to compute summary statistics for a dataset using Python. In addition, since Python is an open-source language, there was no cost to use the tools at hand or to write the code in a Jupyter notebook.

The pizza sales dataset was imported into the notebook using the "Import CSV" line. I used Python's "index()" method to determine an item's location within a list of items. My method of calculating typical summary statistics involved calling each column in the dataset and analyzing it. Using the max and min functions, for example, the maximum and minimum total order prices for pizza are 83.0 and 9.75, respectively. By combining if statements and for loops, I was able to determine that the Thai Chicken Pizza sold the most units (43434.25).

The addition of the else conditional statement enabled the completion of a code line that calculated the money generated by the various pizza sizes offered.

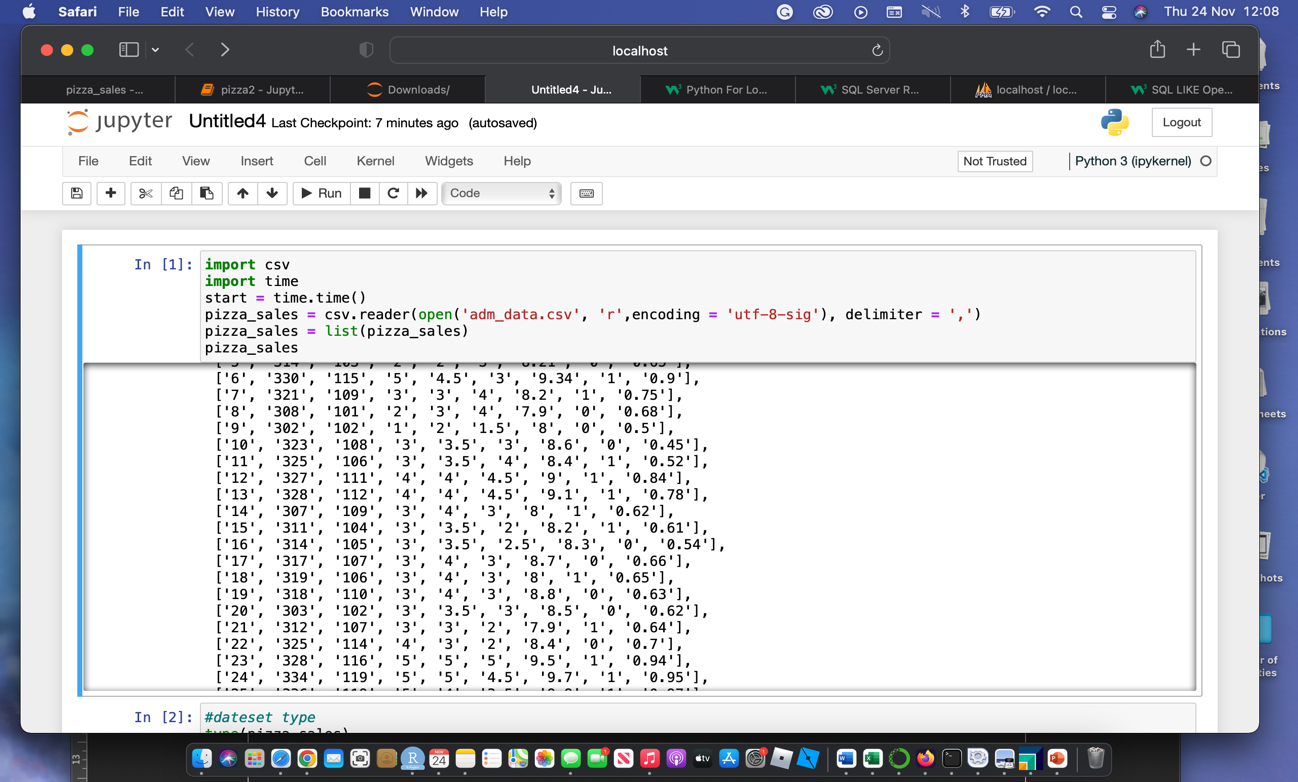
To accomplish Task 2, I learned how to quickly generate data structures by using the Pandas tool. I found that after generating a data frame in Pandas, using the 'import pandas as pd' line of code, I could conduct the same summary statistics, such as obtaining the average pizza sale, the total amount of pizza sold, the best- and least-sold pizza types, etc., with fewer lines of code.

The Pandas function made analyzing and comprehending the data from the second task very simple. In contrast to Task 1, I didn’t have to use "for loops" or "if" statements to generate summary statistics. Most of the time, I only needed a single line of code to achieve the desired result.

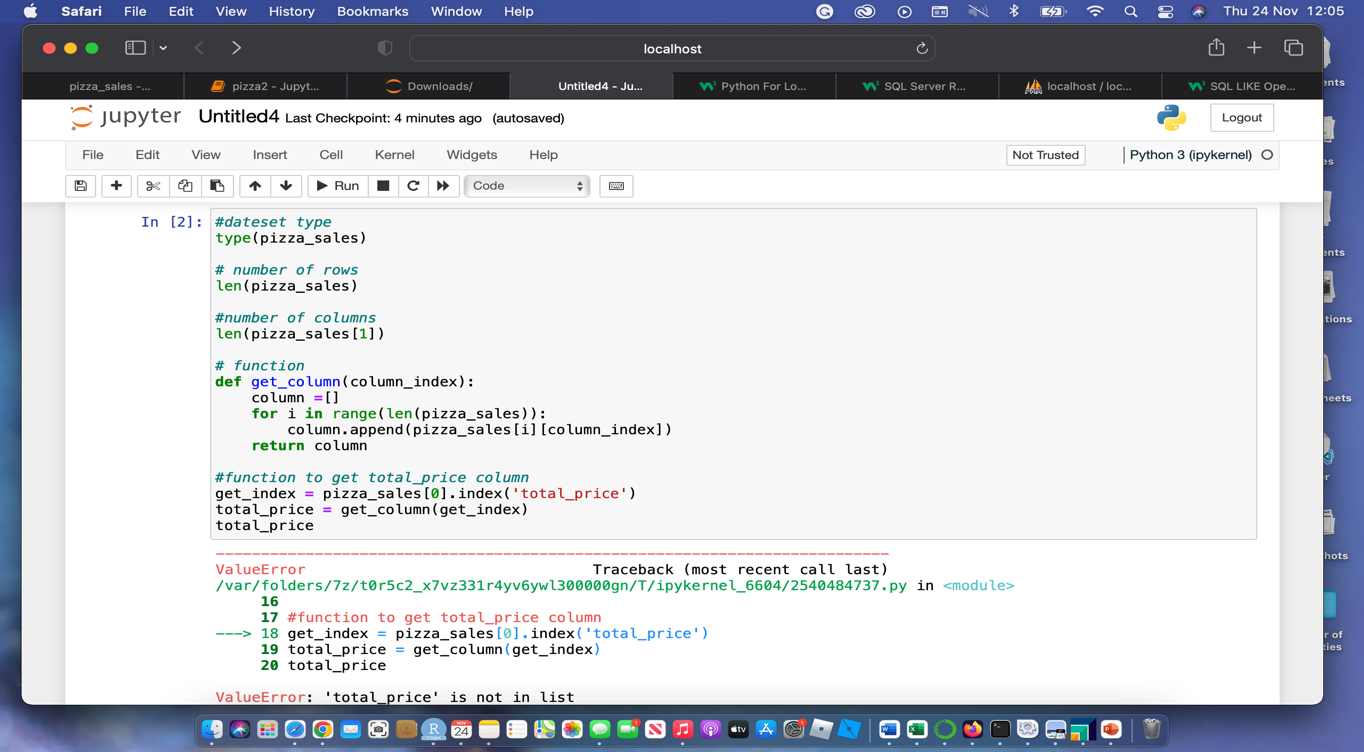
Finally, I used the import time module to compare the execution times of the code from Tasks 1 and 2 (Pandas) using the same datasets.

As Task 2 was analyzed using the Pandas function, it took less time to complete than Task 1. This serves as another evidence that Pandas make data analysis efforts more efficient and produce better outcomes.

**Appendix**

**Fig. 1**

**Fig. 2**

****

**Table 1**

|  |  |  |
| --- | --- | --- |
|  | **Normal dataset size** | **Increased dataset size** |
| **Task 1 (Duration)** | 1.492 | 2.189 |
| **Task 2 (Duration)** | 0.595 | 0.692 |