

SIT112 - TASK COMPLETION REPORT

Important note 1: Please **do not** include Python code in this report. It would be acceptable though to make brief references to function names or different parts of the task's notebook when needed.

Important note 2: Please **do not** leave any response field empty; if not applicable, enter NA.

1. TASK SUMMARY

TASK NAME (ABBREVIATION)	YOUR NAME (STUDENT ID)	YOUR DEAKIN EMAIL
Credit Task: C3	Jason Vellucci SID: 221437402	jvellucci@deakin.edu.au

2. TASK DESCRIPTION

ITEM	RESPONSE
What was the objective(s) of the task?	The objective of this task was to further clean, prepare, and analyze the provided avocado sales data.
What kind of data did you work with?	One data set representing the sales statistics of avocados of different types and for different regions. This data was provided as a .csv file.
Briefly describe the data science task you worked on.	The data was cleaned (filtered), prepared (grouped by year, region, type), had aggregate operations performed on it, was displayed in a categorical (bar) plot, and ultimately had the coefficient of variation (the ratio of std deviation to the mean) calculated for it.

3. TECHNICAL SKILLS

ITEM	RESPONSE
What technical skills did you use during the task?	Data cleaning (filtering), data preparation/pre-processing (grouping, filtering, aggregate operations), categorical graphing (bar plot), library-specific skills (Pandas).
list any challenges or obstacles you faced while working on the task and how you overcame them.	The main challenge I faced was my lack of familiarity with some terms, such as the Coefficient of Variation (CV). I overcame this by using CoPilot and web-resources to learn as I went.

4. DATA CLEANING AND PREPARATION (ENTER NA WHEN NOT APPLICABLE)

ITEM	RESPONSE
What steps did you take to clean and prepare the data?	<p>Cleaning: Filtering out of the 'TotalUS' region was the only cleaning of the dataset that I did. This could arguably fall under 'Preparation'.</p> <p>Prepare: Filtering out the 'TotalUS' region (presumably to prevent overlapping), filtering out records where the year was < 2017, grouping the data by <i>year</i>, <i>region</i>, <i>type</i> fields, performing aggregate operations to dynamically generate columns on the dataset.</p>
Did you encounter any issues with the data during this process? How did you address these issues with the data?	<p>NaN was produced in some records when calculating the CV. This seemed to be due to the CV being calculated in fields with a '0' value. No further action was taken</p>

5. DATA ANALYSIS (ENTER NA WHEN NOT APPLICABLE)

ITEM	RESPONSE
How did you analyze the data?	By calculating the Coefficient of Variation (CV) of avocado sale (for each bag type) and generating a categorical (bar) plot.
Did you use any visualization techniques to better understand the data?	Yes, I generated a categorical (bar) plot which displayed the regions with the top 5 volumes of avocados sold. Each region showed the <i>mean</i> , <i>median</i> , and <i>std</i> on the y-axis.
What insights did you gain from this analysis?	Analysis of the bar plot showed that the 'Northeast', 'West', 'GreatLakes', 'California' and 'Midsouth' regions had the highest volumes of avocado sold out of all regions in the data set. In particular, the Northeast and West regions showed a higher std to the mean for Total Volume of avocados sold.

6. BASIC REQUIREMENTS FOR THE TASK

ITEM	RESPONSE (YES/NO)
Are you confident to execute the Python code in this task and explain the output?	Yes
Are you confident to explain what each line of code does and how it contributes to the solution(s)?	Yes
Are you confident to rewrite or modify the code after completing this task? <ul style="list-style-type: none"> For pass tasks: with guidance, no time limit. For credit tasks: with limited guidance, no time limit. For distinction tasks: independently, no time limit. For high distinction tasks: independently, in a limited time. 	Yes

7. CODE ATTACHMENT (NOT APPLICABLE TO THE PASS TASKS – ENETER NA)

ITEM	RESPONSE (YES/NO/NA)
Have you attached the notebook file that contains your solutions (Python code) for this task?	YES
Have you executed all the cells in your attached notebook and ensured there is no error? <i>Please note your submission will not be flagged as complete if your attached notebook contains any error.</i>	YES

8. VIDEO ATTACHMENT (NOT APPLICABLE TO THE PASS/CREDIT TASKS: NA)

ITEM	RESPONSE (THE VIDEO LINK/NA)
Provide the link to the video recording that presents your completed task. This is only for Distinction and High Distinction tasks. Enter NA for Pass/Credit tasks.	NA

9. ACKNOWLEDGEMENT

BY SUBMITTING THIS REPORT, I ACKNOWLEDGE THAT:

- MY RESPONSES ARE ACCURATE AND ARE MY OWN WORDS.
- I HAVE MET ALL THE BASIC REQUIREMENTS OF THE TASK (LISTED IN SECTION 6).
- I HAVE READ AND FULLY UNDERSTOOD THE ASSESSMENT GUIDELINE OF THE UNIT.
- THIS REPORT DOES NOT EXCEED 3 PAGES.
- THIS REPORT DOES NOT INCLUDE CODE EXCEPT BRIEF REFERENCES TO FUNCTION NAMES OR DIFFERENT PARTS OF THE TASK'S NOTEBOOK.
- MY SUBMISSION DOES NOT CONTAIN ANY CREDENTIALS (E.G., PASSWORD, API KEY, ETC) OR PERSONAL INFORMATION.

IMPORTANT NOTE 3: IF YOU HAVE ANSWERED NO TO ANY OF THE QUESTIONS IN SECTIONS 6, PLEASE RECONSIDER SUBMITTING YOUR REPORT; ASK HELP FROM YOUR TUTOR.

ADD YOUR NAME AND SIGNATURE HERE: JASON MARK VELLUCCI