

# 2021 VCE Applied Computing: Data Analytics external assessment report

## General comments

The 2021 VCE Applied Computing: Data Analytics examination was the second year of the *VCE Applied Computing 2020–2024 Study Design*. The 2021 examination consisted of three sections: Section A – Multiple-choice questions, Section B – Short-answer questions and Section C – Case study questions.

In Section A the multiple-choice questions were answered very well. Areas in which students scored highly related to networks, data security, data visualisations and secondary data sources. Areas in which students did not score well were data integrity and data types.

In Section B it was evident from responses to the short-answer questions that students understood the difference between archiving and disposal of files and data types. Responses were also strong in the areas of project management, data collection and security.

Areas for improvement include basic definitions from the study design, such as the difference between validating and verifying data, as well as table normalisation, software functions and testing techniques.

In Section C the writing of criteria for checking the integrity of the collected data (Question 5) was completed poorly; while a criterion does not have to be in the form of a question, for students who struggle with the concept it is the best way to formulate one and the statement must be measurable. A large number of students just wrote one word and could not be awarded a mark for that part. A criterion needs to be more than one word to gain any marks. Also completed poorly was the outlining of functional requirements of the dynamic data visualisation (Question 6a.); students either stated a non-functional requirement or just listed a functional requirement that did not relate to the case study. To gain a mark for each functional requirement it needed to relate to the case study. Students struggled to normalise to 3NF the supplied database (Question 7a.); they were still including the Total Cost in the database or forgetting to indicate the primary and foreign keys.

Other areas that should be addressed include the following.

- When answering a question that asks to justify your response for three marks, the response must involve a comparison of the stated option against the alternative option and why the stated option is preferred (see Section B, Question 2a.).
- When quoting legislation, such as one of the acts, students must include the correct year to receive marks for that component of the question, e.g. *Privacy Act 1988*.
- Questions that ask for two or more items to be addressed with ‘and’ joining them, must have both individual items addressed in the answer or the student will not receive full marks. See Section C, Questions 5 and 11.
- Students need to be familiar with the Applied Computing study design and the Software tools and functions document from the start of each year.

## Specific information

Note: This report provides sample answers or an indication of what answers may have included. Unless otherwise stated, these are not intended to be exemplary or complete responses.

Student responses reproduced in this report have not been corrected for grammar, spelling or factual information.

The statistics in this report may be subject to rounding resulting in a total more or less than 100 per cent.

## Section A – Multiple-choice questions

The following table indicates the percentage of students who chose each option.

Question	Correct answer	% A	% B	% C	% D	Comments
1	A	92	7	1	0	
2	A	34	1	60	5	Only use a numeric data type in a database if you are planning on using it in a calculation. A postcode wouldn't be used in a calculation, so C is incorrect
3	C	2	1	90	6	
4	D	0	14	2	84	
5	B	8	83	4	6	
6	C	6	8	80	6	
7	A	52	18	20	10	A is correct as the customer feedback has been grouped into like topics, which is coding the data. The other three options do not relate to data manipulation.
8	A	52	32	5	11	The most appropriate is a storyboard. A is the correct answer. A storyboard shows the relationship with the components of the dynamic data visualisation
9	A	80	3	15	2	
10	C	5	21	67	7	
11	B	46	42	3	8	From the study design: integrity of data includes accuracy, authenticity, correctness, reasonableness, relevance and timeliness. Answer A includes completeness, which is not included. The correct answer is B. Students need to be familiar with this particular point of key knowledge.
12	A	87	4	3	7	
13	C	6	26	66	2	
14	D	0	0	0	99	
15	B	15	65	15	5	

Question	Correct answer	% A	% B	% C	% D	Comments
16	C	33	3	58	6	C is the correct answer, the other answers do not relate to the calculation of the percentages.
17	D	7	4	3	85	
18	A	62	11	1	25	
19	A	83	10	5	2	
20	C	5	21	65	9	

## Section B

### Question 1a.

Mark	0	1	Average
%	49	51	0.5

Students did not answer this question well. Acceptable answers were mock-up, mock-up diagram, layout, layout diagram and annotated diagram.

### Question 1b.

Mark	0	1	Average
%	45	55	0.6

Acceptable design principles for the appearance of the input form were alignment, balance, contrast, space and image use.

### Question 1c.

Mark	0	1	2	Average
%	65	30	5	0.4

Students were asked to describe two techniques for cleaning data. A lot of students were unable to find two answers.

Techniques that were awarded marks described the need to remove measurement in the height (cm) and handspan (cm) cells and place this in the field heading or check that common units of measurements have been entered in the height field (i.e. if an entry was 2.01 m, it should be changed to 201 cm and recorded as 201, so it is the same as the other entry).

The following is an example of a high-scoring response.

*Converting all numerical data to be the same unit of measurement i.e. The height value that is in meters should be changed into centimeters. The units of measurement should also be removed from the cell and instead placed in the heading in brackets e.g. data name (cm) for height and handspan.*

## Question 2a.

Mark	0	1	2	3	Average
%	13	13	45	30	1.9

To be awarded full marks, students needed to select the infographic, state why it was selected and then state why the dynamic visualisation was not suitable.

The following is an example of a high-scoring response.

*An infographic. The manager is looking for a static poster, that contains text, statistics, images and charts. An infographic can do this, therefore a dynamic data visualisation that is interactive and changes in real time is not necessary.*

## Question 2b.

Mark	0	1	2	Average
%	63	26	11	0.5

The study design refers to a format as including images, graphs, tables and text. A large number of students were not able to link these key words to their response and as a result were not able to gain any marks for this section of the question.

Students who gained a mark for a convention linked their responses to the definition of a convention from the study design. This included:

- title/heading using bold or larger font size than the rest of the text
- text/font style – a well-chosen typeface such as Arial, Times New Roman, Courier New or a slab serif typeface
- a legend when using charts
- appropriate colour and contrast such as shades best used for background, avoid using yellow font on a white background, limit the number of colours etc.

The following is an example of a high-scoring response.

*Format: Charts and graphs can be used as they make it more engaging and readable as easy to understand.*

*Convention: For conventions, big bold headings and readable text styles can be used.*

## Question 3

Mark	0	1	2	3	Average
%	15	31	35	19	1.6

To gain marks for the legal requirement section, students needed to state the *Privacy Act 1988* (must include the correct year) or one of the Australian Privacy Principles related to the Privacy Act.

Students were then required to describe the legal requirement including what the manufacturer needed to do to meet this requirement.

The following is an example of a high-scoring response.

*Comply with the Privacy Act of 1988: As a car manufacturer, likely turnover of 3 million dollars, they will have to comply with the Privacy Act 1988 as they are based in Australia. An example would be that they have to disclose their data collection to customers.*

## Question 4a.

Mark	0	1	2	Average
%	18	51	31	1.2

To obtain marks, students needed to explain why it is important to generate more than one design idea. The most common responses were:

- the first design idea is not necessarily the best one
- a different design may be better suited in meeting the client's demands
- designs can be a personal choice; providing a range allows the client to choose what would suit them the best.

## Question 4b.

Mark	0	1	2	Average
%	77	17	6	0.3

A large number of students misread the question and simply stated evaluation criteria. The question was asking how to develop the evaluation criteria and therefore these responses were not awarded any marks; very few students were awarded full marks for this question.

Responses that scored highly mentioned one of the following.

- Consider the functional and non-functional requirements of the solution.
- Look at design principles to formulate criterion links to one of the principles.
- Use the measures of effectiveness to formulate a question or statement.

The following is an example of a high-scoring response.

*Design evaluation criteria can be developed based on the functional and non functional criteria, such as using the design principles for non functional aspect. These help determine a preferred design by giving a somewhat objective method of selecting a design, reducing indecision or subjective evaluation.*

## Question 5a.

Mark	0	1	2	Average
%	9	35	56	1.5

To gain full marks, students needed to explain the difference between archiving files and disposing of files. This included an explanation of what each was.

A high-scoring response needed to include something similar to the following.

- Archiving places files in medium- to long-term storage. They are usually compressed to preserve storage space and can be accessed at a later date using archive software.
- Disposal involves placing a file in the recycle or trash, which marks the file to be deleted. The deletion takes place once the recycle/trash is emptied. This frees up more space than archiving as it is permanent deletion of the files.

## Question 5b.

Mark	0	1	2	Average
%	20	46	34	1.2

Many students were unable to gain full marks as they did not fully answer the questions. For example, a common advantage response was that data can be accessed from any device. Students needed to add that they needed an internet connection.

Students answered the disadvantage section very well.

### Advantages

- Cost – purchasing physical storage can be expensive. Without the need for hardware, cloud storage is exceptionally cheaper per GB than using external drives.
- Accessibility – using the cloud for storage gives you access to your files from anywhere that has an internet connection.
- Recovery – in the event of a hard drive failure or other hardware malfunction, you can access your files on the cloud. It acts as a backup solution for your local storage on physical drives.
- Syncing and updating – when you are working with cloud storage, every time you make changes to a file it will be synced and updated across all of your devices.
- Size – can easily obtain large amounts of storage space online.
- 24/7 access using any device.

### Disadvantages

- Internet connection – cloud-based storage is dependent on having an internet connection. If you are on a slow network, you may have issues accessing your storage. In the event you find yourself somewhere without internet, you won't be able to access your files.
- Cost – there are additional costs for uploading and downloading files from the cloud. These can quickly add up if you are trying to access lots of files often.
- Support for cloud storage isn't the best, especially if you are using a free version of a cloud provider. Many providers refer you to a knowledge base or FAQs.
- Privacy/Security – when you use a cloud provider, your data is no longer on your physical storage. Who is responsible for making sure that data is secure? That's a grey area that is still being figured out.
- Less control over your data, including backup.
- Security – greater chance of a security breach.
- Deletion of data – do you know if your data has been deleted or not?

## Section C

### Question 1

Mark	0	1	2	Average
%	45	20	35	0.9

After reading the case study, students were asked to write a research question that would assist the Vice President of Sales to present the findings to the board.

Students need to understand that a good research question must be clear, focused, concise, complex and arguable, therefore the question should not be able to be answered with a yes or no.

The question should have started with 'What', 'Why' or 'How' and related to the scenario.

For example:

- How can XPLUSMISC increase their revenue as there is a decline in fuel sales over the last few years?
- What services will encourage our customers to shop with us?

The following is an example of a high-scoring response.

*Why has there been a decrease in fuel sales over the last 10 years at XPLUSMISC?*

## Question 2

Mark	0	1	2	Average
%	25	39	36	1.1

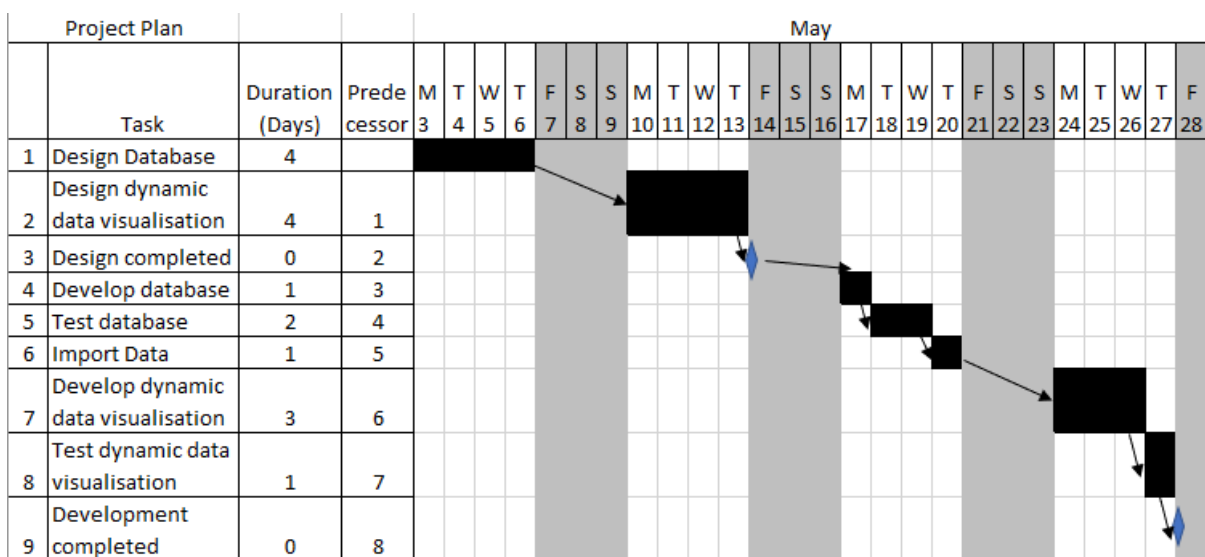
Students were required to provide one reason why XPLUSMISC would want to acquire customer data through the use of a new loyalty card.

The most common responses that were awarded marks were:

- support customer loyalty by providing targeted offers based on purchases
- can develop profile of customers based on suburb/state/gender and improve target marketing strategy
- obtain an understanding of customer purchase trends
- tailor offers to specific customers for special occasions
- basis of a mailing list for marketing campaigns
- use of contact details for marketing campaigns
- discounts increase purchases and can increase their customer base or sample size.

## Question 3

Marks	0	1	2	3	4	5	6	Average
%	5	9	10	20	24	13	18	3.6



Marks were awarded for:

- correct start date
- correct finish date
- correct durations
- two correct milestones
- correct dependencies.

To gain full marks students needed to make sure that the milestones were on the line between the two days; a number of students placed the milestones in the middle of the day and therefore did not gain a mark for that component. Students also needed to make sure the dependencies went from the completion of the previous task to the start of the next task. Students who did not do this could not gain a mark for the dependencies.

## Question 4

Mark	0	1	2	3	4	Average
%	13	13	29	21	25	2.4

Method of data collection	Example	Justification
Surveys	<ul style="list-style-type: none"> <li>• Customers fill out paper survey on coffee needs</li> <li>• Customers fill out survey sent by email</li> <li>• Customers fill out survey that is mailed out</li> <li>• Marketing staff phone customers</li> <li>• Can use a Likert scale for particular questions</li> </ul>	<ul style="list-style-type: none"> <li>• Convenient way to collect a lot of data quickly</li> <li>• Can get personal responses more easily</li> </ul>
Observation	<ul style="list-style-type: none"> <li>• Staff record items purchased by customer</li> <li>• Staff count stock items after each day</li> <li>• Count visitors to the store</li> </ul>	<ul style="list-style-type: none"> <li>• Can be very time consuming but accurate</li> <li>• More detailed responses as to why something is required</li> </ul>

Marks were awarded for each example of a collection method and how suitable that method would be.



The following is an example of a high-scoring response.

Method of data collection	Example	Justification
Surveys	<i>Kamal could conduct an online survey with questions about their spending, liking for products in a Likert scale, Boolean questions on if they would like a barista and a feedback question on what else they would like to see,</i>	<i>Conduct a survey would let them collect a large amount of data in a small period of time and would be very efficient in terms of time money and human sources</i>
Observation	<i>Karmal could observe the spending of customers earnings over a period of time at XPLUSMISC. They would have to observe people Behaviour with products at stores as well as data about daily earnings in categories.</i>	<i>Observations would provide rich and high quality of data, however Kamal only has 12 months to collect and manipulate data and present it to the board. This would also be less efficient in forms time, money and human resources as it will be expensive in all aspects.</i>

## Question 5

Mark	0	1	2	3	4	Average
%	47	19	22	4	7	1.1

The criteria must have been a question or statement (not a single word) and must have related to the case study.

When identifying the criteria, they needed to relate to accuracy, authenticity, correctness, reasonableness, relevance or timeliness.

The following is an example of a high-scoring response.

*Is the collected data accurate?*

*It is important that the data the Kamal collected is accurate as otherwise any results produced for the data will be misrepresented and will not reflect the truth*

*Is all of the data authentic?*

*The data needs to come from reputable and reliable sources, otherwise the decisions made based off the data may be incorrect and could damage the integrity of the data.*

## Question 6a.

Mark	0	1	2	Average
%	72	21	6	60.4

While students were generally able to indicate two functional requirements, most were not able to link them to the data visualisation.

Responses that gained marks included:

- ability to navigate between the individual screens
- displays the correct proportions in a chart representing values
- displays screen without time lag
- displays top five requested items
- displays answer to barista question
- contains navigational elements.

## Question 6b.

Mark	0	1	2	Average
%	57	31	12	0.6

Responses needed to relate to the quality of attributes of the solution.

Responses that gained marks were similar to:

- have contrasting colours for people with visual impairments
- needs to be easy to use
- should be visually appealing.

## Question 7a.

Mark	0	1	2	3	Average
%	74	15	8	2	0.4

Reward_ID	Transaction_ID	ProductPurchased_ID
Surname	DateOfPurchase	ProductName
FirstName	Store_ID	ProductCost
Postcode	Time	
Mobile_Number	ProductQuantity	
NumberOfPoints	ProductPurchased_ID	
	Reward_ID	

One mark was awarded for each of the following:

- correct fields in each table
- identifying the three primary keys
- ProductPurchased\_ID and Reward\_ID in Transaction table.

Customer\_ID or Reward\_ID were accepted.

## Question 7b.

Mark	0	1	2	3	Average
%	8	13	36	43	2.2

Responses that did not score well usually indicated that the Mobile\_Number data type was numeric.

Students need to remember that only fields used in a calculation need to be numeric; mobile number is not one of them.

Acceptable answers were:

- Reward\_ID: Text/String/Character
- Mobile\_Number: Text/String/Character
- NumberOfPoints: Integer/Numeric

## Question 8

Mark	0	1	2	3	4	Average
%	49	5	13	6	26	1.6

It was clear from the responses that students were not familiar with the different types of design tools, with a variety of responses that did not state an actual design tool.

Acceptable answers were:

- storyboards
- layout diagrams
- annotated diagrams
- mock-ups.

## Question 9

Mark	0	1	2	3	Average
%	21	35	30	13	1.4

A set of procedures is a number of steps or separate ideas that relate to a particular item. To gain full marks, students were required to propose a set of procedures to manage the security of the files.

Acceptable answers were:

- encrypt the data
- set up access control procedures such as usernames and passwords to access files
- set up access restrictions – only certain XPRESSPLUS staff can access files
- set up policies around the use and access to the data
- de-identify data in files
- reduce the number of copies of the data (this is the most important – spreadsheets are generally why corporate data leaks)
- set up security education training and awareness around this application
- ensure that appropriate physical security controls to where the data is used and stored are implemented
- ensure that appropriate software security controls around the data are implemented (firewalls etc.)

The following is an example of a high-scoring response.

*Kamal should implement software security measures such as username and password to minimise unauthorised access. He should then implement physical security controls such as locking the door of the room the database files in. He should also implement a backup strategy at the daily incremental and weekly full back ups to ensure that he can recover the data in the event of data breach or loss.*

## Question 10a.

Mark	0	1	2	Average
%	43	28	28	0.9

To gain full marks students needed to describe two suitable techniques for navigating through the dynamic data visualisation. The responses needed to relate to the image that was presented as part of the question.

Responses that gained marks used the following techniques.

- Click on the states to show information on that state.
- Click on each of the charts to show more detailed information on each (e.g. if they want to see more detail on items, click on the item chart to then show item by region or store, or show more than the top five).
- Use buttons to go between screens.
- Use the navigation buttons to go through state by state.
- Use of the state hotspots.
- Have a zoom button to view information more closely.
- Create dropdown list of the states.
- Use of hyperlinks.

The following is an example of a high-scoring response.

*The user can click on the arrows which bring them to another page*

*The user can use links to navigate to another graph in the visualisation*

## Question 10b.

Mark	0	1	2	3	4	Average
%	56	7	23	4	11	1.1

Students were expected to draw on their knowledge of the software function from the [Data Analytics Software tools and functions document](#) (published each year on the study page) and only responses that related to this were accepted.

Responses that gained marks used the following functions:

- perform calculations
- description – calculation will need to be performed to determine the top five products (by %)
- formatting the dynamic data visualisation charts/graphs
- creating hyperlinks and other navigations
- sorting the data
- filtering the data
- creating queries in the database to allow data to be graphed
- deleting unwanted data
- use of images
- conditional formatting.

The following are examples of high-scoring responses.

*Charting/Graphing: Kamal should use the graphing software to produce charts to use in the dynamic data visualisation*

*Calculations: By calculating the percentage of respondents interested in a barista / their spending habits, the proportion of answers can be visualised.*

*Sorting: the top 5 requested items can be sorted from the data based on a descending order.*

## Question 11a.

Mark	0	1	2	Average
%	71	25	3	0.3

To gain full marks students needed to explain the difference between validating and verifying data.

Students needed to give a statement about each item, explaining what they were. The study design (refer to the 'Terms used in this study') clearly outlines what each item is and therefore the expected responses needed to have that detail.

Data validation is the checking of data for its reasonableness and completeness. Validation of data includes: existence or presence checking, which verifies that a required field has a value entered and is not empty or blank; range checking, which involves ensuring that data entered falls within a certain range; and type checking, which confirms that data entered is of a particular type (must include both reasonableness and completeness).

Verifying is checking of data after data entry. This can include proofreading to ensure that the data entered matches the source data.

## Question 11b.

Mark	0	1	2	Average
%	85	9	6	0.2

Possible responses were:

- Is the amount entered for ProductQuantity a valid number in a reasonable range?
- Surname has a required value
- ProductQuantity has a range check of 0 to 100

Responses needed to be an example of an existence check and relate to the case study.

## Question 11c.

Mark	0	1	Average
%	65	35	0.4

Possible responses were:

- Check/proofread for duplicate records
- Check/proofread to determine that the address entered for the customer is correct
- Ensure data entered into the data visualisation matches source data

## Question 12

Mark	0	1	2	Average
%	61	21	18	0.6

To gain marks the students were required to outline one suitable testing technique for navigating through the dynamic data visualisation.

This included identifying the navigational element from the data visualisation and then outlining how to test that the element would work.

The following is an example of a high-scoring response.

*Kamal should produce test data for navigating the visualisation and determine the expected result for these test data. For example, Kamel could test the right arrow. The test data could be left click, right click and not clicking on it. The expected result would be move to the next page, nothing and nothing respectively.*

## Question 13

Mark	0	1	2	Average
%	73	19	8	0.4

Students were expected to indicate the timeframe during which the evaluation should occur (3–6 months) and how they would go about the evaluation (interview, survey or direct observations).

The following is an example of a high-scoring response.

*3-6 months after the implementation of the visualisation, Kamal should conduct interviews with the users of the visualisation. This data could be used by Kamal to determine to what extent the data visualisation meets both the functional and non-functional requirements.*

## Question 14

Mark	0	1	2	3	Average
%	22	30	32	17	1.5

Possible responses were:

Type of threat: Event-based (natural disaster)

Strategy: A good backup procedure or use a UPS/generator

The following is an example of a high-scoring response.

*Event based treat*

*To protect the organization from this threat, XPLUSMISC could apply a data backup strategy across all its stores, who should be assigned to take weekly full backups and daily incremental backups, and stores this off-site on the cloud, this way a power surge will not lead to permanent customer transaction data loss since a copy of data is located off-site to be retrieved.*

## Question 15

Mark	0	1	2	3	4	Average
%	11	6	16	31	36	2.8

Students were asked to complete the table provided by identifying two types of physical security controls and two types of software security controls and then describe how each type of control will improve XPLUSMISC's current data and information security practices.

The most common acceptable controls were:

- Physical security controls
  - Locks on doors
  - Barrier techniques
  - Cameras/CCTV
  - Biometrics
  - Alarms
  - Security guards
- Software security controls
  - Access restrictions
  - Usernames and passwords
  - Access logs
  - Encryption
  - Password protection of files.

The most common acceptable descriptions were:

- Physical security controls
  - This will prevent unauthorised users from accessing the server room.
  - This will prevent customers from coming behind the counter and accessing the computer.
- Software security controls
  - This will prevent unauthorised users from gaining access to the computers.
  - This will prevent staff that don't have authorised access from gaining access to customer files.

Marks were awarded for identifying the control and a suitable description that matches the control. The control needed to match the content in the questions. No marks were awarded if there was no description or either the control or description was incomplete.

The following is an example of a high-scoring response.

*Physical security controls*

- *Physical Locks: A physical lock will ensure no unauthorised access into the server room will occur, resulting in less of a chance of data loss.*
- *CCTV Cameras: Will deter customers from even attempting to get behind counters and access computers*

*Software security controls*

- *Username and passwords: Username and passwords will allow an employee's personal information and business details to stay secure from customer and other employees.*
- *Access restrictions: These will ensure that employees at XPLUSMISC will only be able to access files that are granted to them meaning they will not be able to access files if unauthorised.*

## Question 16

Mark	0	1	2	3	Average
%	20	31	27	21	1.5

A recommended backup strategy could include:

- Frequency (daily, weekly)
- Media (tape, SSD, cloud)
- Location (cloud, offsite) Type (full, any type of partial – incremental, differential).

To gain full marks, students needed to recommend three of the four responses above.

The following is an example of a high-scoring response.

*XPLUSMISC stores should complete a full back up once a week, with incremental back ups daily. These back ups should be stored on an SSD (solid state Drive) and should be locked securely in the store's office. In addition, a copy of the backup should be located off-site. The store could use a wifi connection to store a copy of the back up in a cloud service in addition to the SSD onsite.*