Return Data‏

MegaStore is a large retail outlet, selling everything from groceries to auto parts. They are looking into customer behaviour and trying to understand which purchases end up being returned in order to reduce future returns and refunds.   
  
You are provided with information on items purchased at one of MegaStore’s shops across several days and if a given item was returned by the end of the store’s 14-day return period.   
  
Your task is to analyze the data and answer the following questions.   
The data can be found in returns\_data.csv ﬁle (download below).   
Features

* time\_of\_purchase – timestamp of purchase
* club\_membership – was the client a member of store’s club
* checkout\_type – (regular/self-checkout)
* payment\_method – store code for payment method (credit card, gift card, store card etc)
* product\_code – alphanumeric code indicating brand, type and individual product
* department – department product belongs to
* items\_in\_purchase – total items in same purchase
* cost – item cost at time of purchase
* cashier – personal code of cashier performing the sale
* return – was the item returned by the end of the returns period? (0-no return / 1-return)

Review the data and answer the following questions:   
Q1. What fraction of items end up being returned?  
Q2. What fraction of items costing 1000 or more end up being returned?   
Q3. Which of the following departments has the highest ratio of returns?

* A. Books
* B. Fashion
* C. Grocery
* D. Small appliances
* E. Outdoor furniture
* F. Toys

Q4. Which of the following departments has the highest number of returns?

* A. Books
* B. Fashion
* C. Grocery
* D. Small appliances
* E. Outdoor furniture
* F. Toys

**Part two**  
  
  
In this question you are provided again with the dataset of item returns from the previous question . The data can be found in returns\_data.csv ﬁle, which you used in the previous question.   
  
  
MegaStore is a large retail outlet, selling everything from groceries to auto parts. They are looking into customer behaviour and trying to understand which purchases end up being returned in order to reduce future returns and refunds.   
You are provided with a dataset representing items bought at the store over a period of several days, and whether any given item was returned by the end of the return period.   
Explore and analyse the data provided, and try to understand the parameters indicating if a given item purchase is likely to end up in a return or not.   
  
You are provided with a list of purchases done a week later, and are asked to predict which ones are most likely to be returned before the end of the return period  
  
Input format  
returns\_test.csv (you should download it below) Output format  
returns\_test.csv given in this question with column return ﬁlled with 0/1 predictions. Note: Do not to change columns order in the table.