Lead Data Scientist - Take Home Tasks

We have provided you with the events tracking data from our site. ([download from here](#))

Please complete the following tasks:
1. **Analyze the given data**
   a. Prepare your assessment of the given data
   b. Develop a suite of 3-5 metrics that can help business understand the performance of the user funnel
   c. Describe the metrics identified above along with their strengths and weaknesses
2. **Generate insights and design experiments**
   a. Create a framework to generate insights from the data and provide 2-3 ideas to the product/engineering team that can drive order completions
   b. Design experiments with the ideas identified above. Put together a *standard operating procedure* that will help product teams take a sound data-driven decision. Please highlight the common pitfalls that should be avoided when analyzing experiments' results.
3. **Repeat Customers**
   a. Help the marketing team understand who are our repeat customers
   b. Help them identify which first time buyers are potential repeat customers

Please feel free to make and state any reasonable assumptions; state the scenarios where your assumptions might break. Please feel free to simplify your approach and elaborate on how the approach can be expanded upon.

Please send back your codes, models, and references (any useful blogs or StackOverflow answers, for example, that have helped you with the tasks). Python would be the preferred programming language, however, you can use any non-proprietary programming language and associated libraries.

A funnel is defined as:

New Session → Page View (eg. collection page) → Product View → Add to Cart → Start Checkout (multiple steps) → Order Completed

Funnel Events:

collection_view - viewed collection page
product_view - viewed a product
add_to_cart - added an item to the cart
checkout_page_view_contact_information - checkout step 1 (checkout initiated)
checkout_page_view_shipping_method - checkout step 2 to select a shipping method (optional when people select express checkout like PayPal)
checkout_page_view_payment_method - checkout step 3 to select payment method
checkout_page_view_processing - checkout payment processing step
checkout_page_view_stock_problems - checkout failure due to stock issue
checkout_page_view_review - checkout step to review the order in case of express shipping
order_complete - order successfully placed

Columns in the events tracking data:

row_id
event_id
event_type - the type of the event as listed above
event_time - in UTC
lb_user_id - a unique user id tied to a browser
email - obfuscated email of the user
product_id
product_category
product_type
order_id
collection_name
session_id
ip
active_experiments
device
url
device_category
user_agent
user_gender
user_state
traffic_source