Data Exploratory analysis [Cleaning/Basic Analysis of Data]

Data for Analysis:

- uicycles - contains the install-uninstall events for customers
- eventdump - contains event triggered within the app for customers

Introduction:

Customer information contains install-uninstall-reinstall cycles in vertical format [~170k records] [so for a customer 1 lifetime cycle is a single install-uninstall event set or a reintall-uninstall event set] [so if a customer install-uninstalls-reinstalls-uninstalls then he has 2 life cycles within the app which need to be treated separately]

Event logs contain the events triggered within the app by these customers [~1 million records]

Generate the following from the data provided [Analysis questions below are along the lines of CLTV (Customer Lifetime Value) analysis]

- Customer retention trends from their lifetime cycles [frequency chart or histogram plot] [retention is defined as the duration of one install-uninstall cycle, so multiple re-installs have to be treated separately]
- Find out the time of day when the customers are most active [use your own discretion for time of day bucketing] [activity is defined on the basis of events]
- Purchase value buckets [find purchase/checkout events from event logs and parse the 'properties' column to get total value associated and generate a simple bucketed frequency chart/histogram plot]
- Behavior of purchasing and non-purchasing customers [something along the lines of their in-app event frequency in a given install-uninstall cycle]
- Week over Week revenue trends for purchasing customers
- How their purchases are distributed post install? [the number and value of purchases after installing the app in one retention cycle]
- Do they perform purchases in the 2nd, 3rd etc weeks post install? [if their retention cycle is greater than 1 week]
- Is there a steady inflow of revenue for customers with high retention? [Growth can decline but is it still a positive gradient?]
- Any other actionable insights that can be drawn from the given data?