Ghostery Assignment

Product Analyst Intern

# Assignment Instructions

1. Read the scenarios and deliverables below
2. Complete and return the assignment to us within 5 business days.

# **Scenario 1:**

# **Overview**

Imagine you are the Product Analyst for Ghostery. Given the nature of the product, the majority of Ghostery’s current 2.5M monthly active users have a high technical acumen (Figure 1). Additionally, Ghostery saw quite notable growth in its first few years, but recent growth has slowed (Figure 2).

To extend their user base, Ghostery is considering releasing a new product**,** *G by Ghostery*, aimed at a more mainstream audience. As the new Product Analyst, it is your job to:

* Create a model to estimate the potential user base for *G by Ghostery*, and forecast growth rates for the new app over the next 24 months
* Outline indicators for success

# **Deliverables**

Create a product analysis plan for *G by Ghostery* that includes a growth model and key indicators of success. In addition to the outputs of your model, this plan should address the following questions:

1. **What data did you use to estimate these values? Where did you source this data from?**

MY ASSUMPTION IS THAT THE NEW PRODUCT IS VERY SIMILAR TO OUR CURRENT PRODUCT(SOLVES THE SAME PROBLEM). IT MAY BE AN IMPROVEMENT WITH NEW FEATURES BUT PRIMARILY PROVIDES ANTI-TRACKING and ANTI-BLOCKING. THE AIM IS TO CREATE IT SO THAT IT IS ADOPTED BY A BROADER AUDIENCE.

**Analysis of current user base:**

1. **Based on average monthly users by year**

For the simplest analysis, I took monthly active users by year as the input for years from

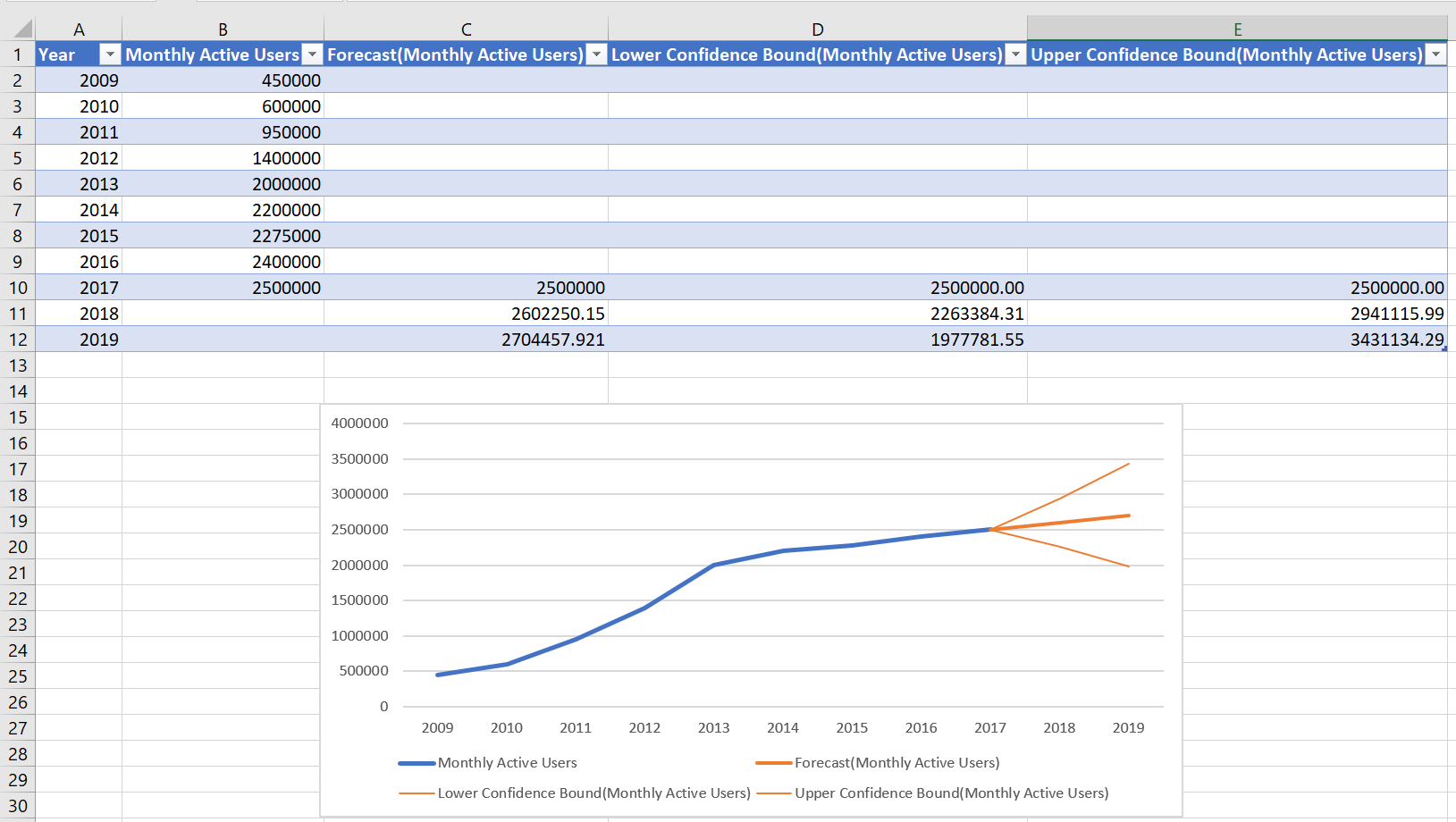
**2009-2017** and found estimated users for 2018 and 2019 based on previous trend.

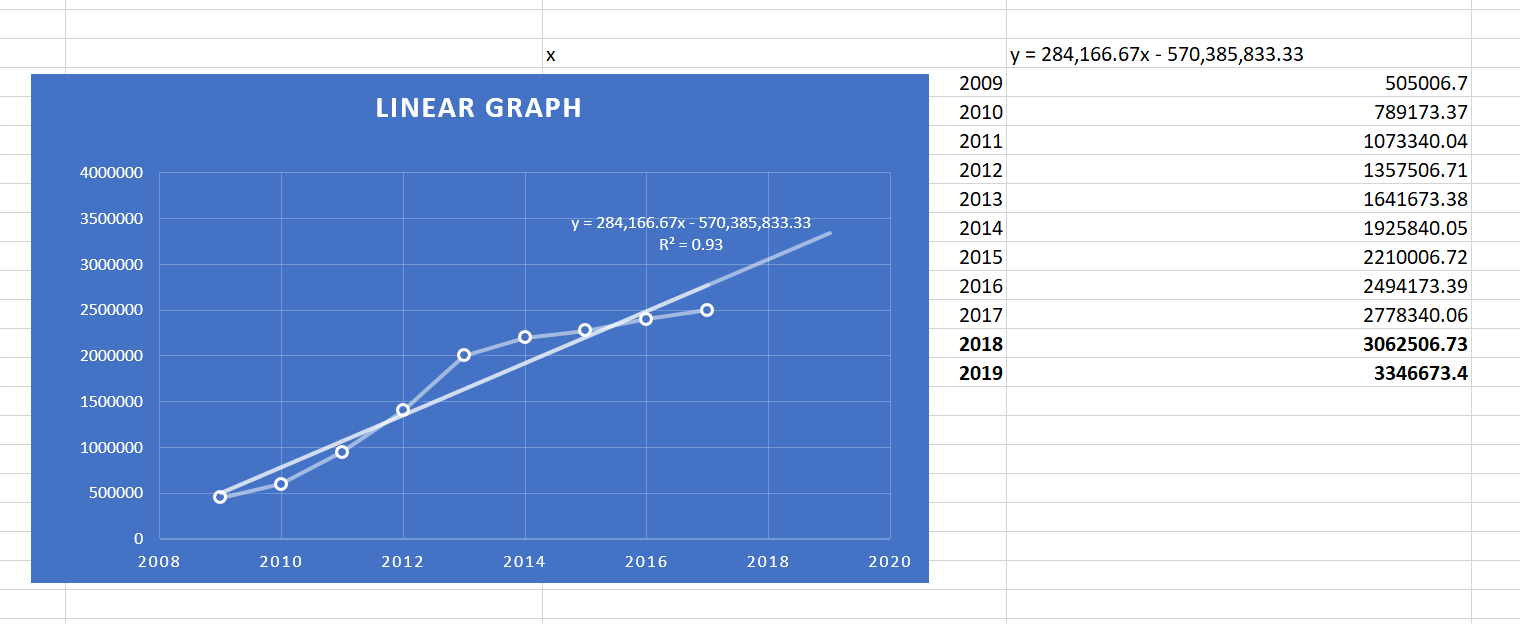
The first chart represents a trend mapping based on linear trends, lower confidence(i.e. when the number of users decreases) and upper confidence(when the number of users increases more than expected).

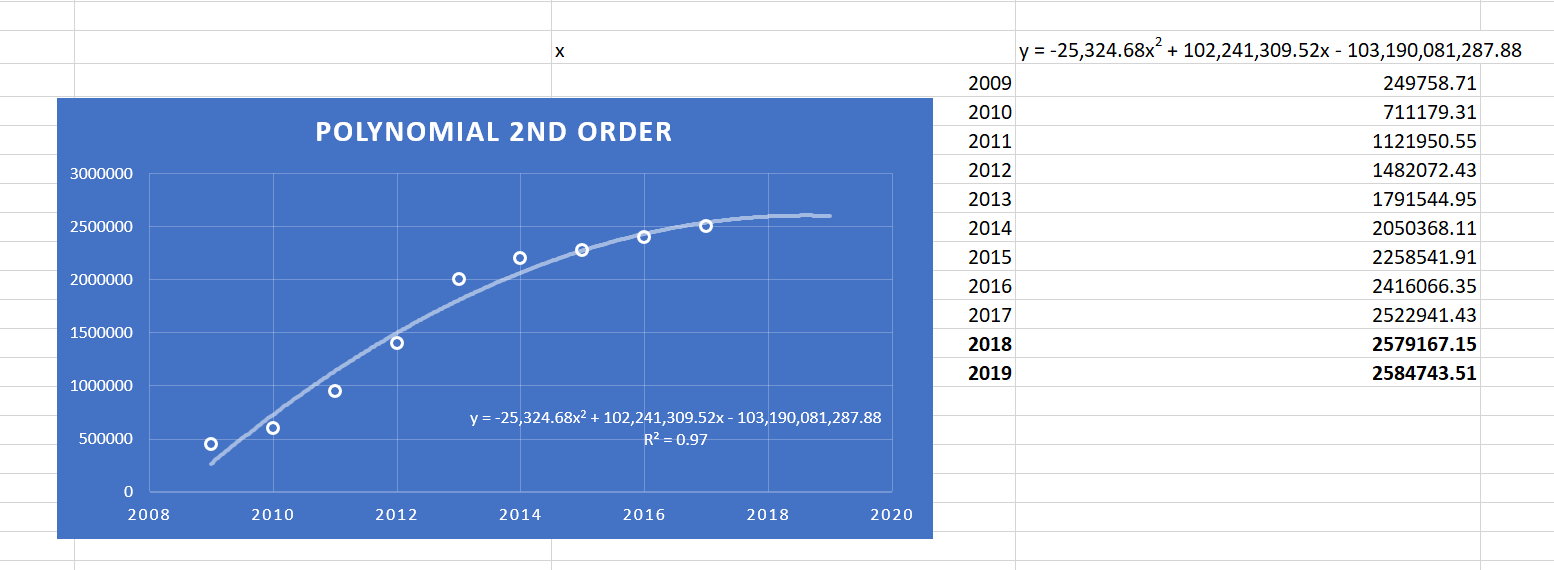
The second chart represents a Linear trend line analysis based on the same data.

Then, this is followed by Polynomial(2nd Order) graphs which give a more flexible(realistic estimate) of how the number of active users per month will vary over 2018-2019.

However, if we go beyond this i.e. to 3rd order or 4th order, the model starts to overfit and thus, only reflects a decline in the user base.







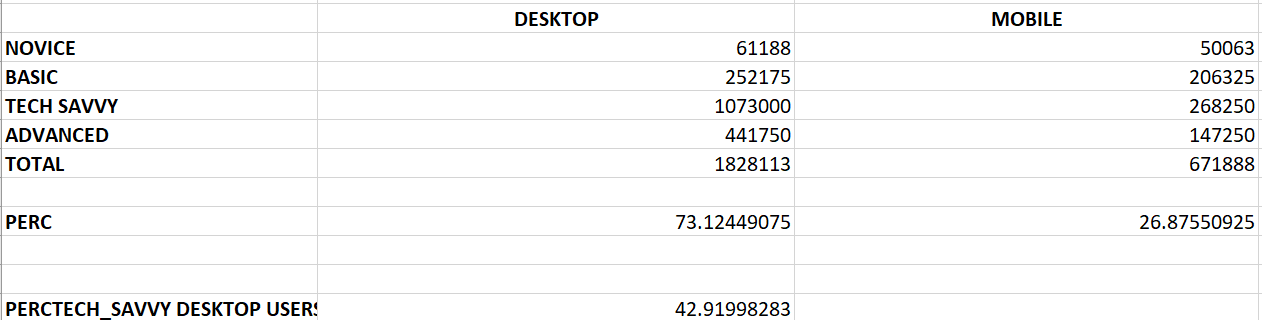
**Adding new feature**

1. **Based on average monthly users by device and technical acumen**

Our matrix for the current month looks like as follows:



Extrapolating from this table, we can decipher that Desktop users comprise of 73.125% of our market whereas Mobile users are just 26.876%. Also, more than half of our user base for this month ~53.65% is Tech-Savvy.



Upon looking more closely at our customer base, we discover that **~43%** of our current customer base is made up of Tech-Savvy Desktop Users.

Thus for the current product analysis, my assumption is that we’ll be looking to serve and capture a larger share of wider market audience.

*Our Product Analysis Plan will be centered around our Desktop users.*

For our Product Analysis, we are done with the first step i.e. narrowing down on who our customer is. We have our Customer Profile (*Persona*) ready.

**Our customer is someone who mostly uses their Desktop to browse on the Internet and does not have to be well versed with technology usage i.e. the product should have certain features that can help a stay-at-home mom, a retired grandpa or a proficient software engineer use the product hassle-free.**

**We will be making a newer, better version of our product based on the same philosophy.**

**What would our Typical Customer do?**

-*Not technically advanced*

*-Likes to use Laptop/Computer mostly*

-*Surfs on the Internet*

*-Opens multiple pages*

*-Does not want to be tracked*

*-Is Time constrained*

*-Does not value ads*

**How would our product address these issues?**

*-One click easy setup*

*-Resilient and Robust*

*-Sharp Back end processing with no lagging*

*-Usable Visual Dashboard*

*-Graphs and Charts for easy illustration*

*-Fast and Efficient*

*-Tracks previous sites visited – Caching*

*-Analyses web usage*

Now let’s look at our analysis subjectively and base our growth model as per that.

**One thing that I’ve learnt from my past experience is that : *Growth Model is about inputs and not outputs!!***

Here, we are taking the assumption that this product is completely raw and will have 2% of our current customer base as early-adopters of the product i.e. **50,000**

Let us take **$1,000,000** as our monthly investment into harnessing the growth of this product.

So for releasing our product, we are assuming that our **Cost per Installation(CPI)** is **$1**

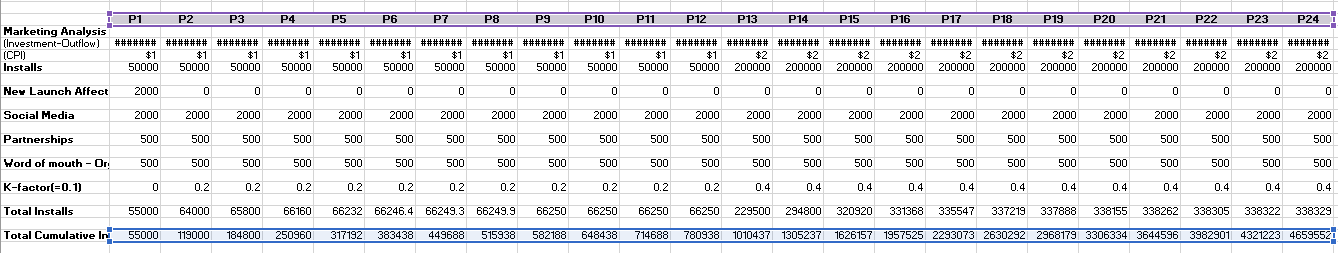
In addition to having the 50,000 customers, we have also taken other customer boosting methods into account:

These are purely assumption based on how product launch is usually handled and we are looking at the model only from the point of view of installations.

For the first 12 months, we have taken following parameters to calculate our installation and therefore, growth.

* **New Launch** Affect gives us an additional 2,000 customers for the first month
* **Social Media** gives us another 2,000
* **Partnerships** give us 500 more customers every month
* **Organic growth** – Word of mouth gives us 500 customers as well
* **K-factor***(growth factor)*-0.2

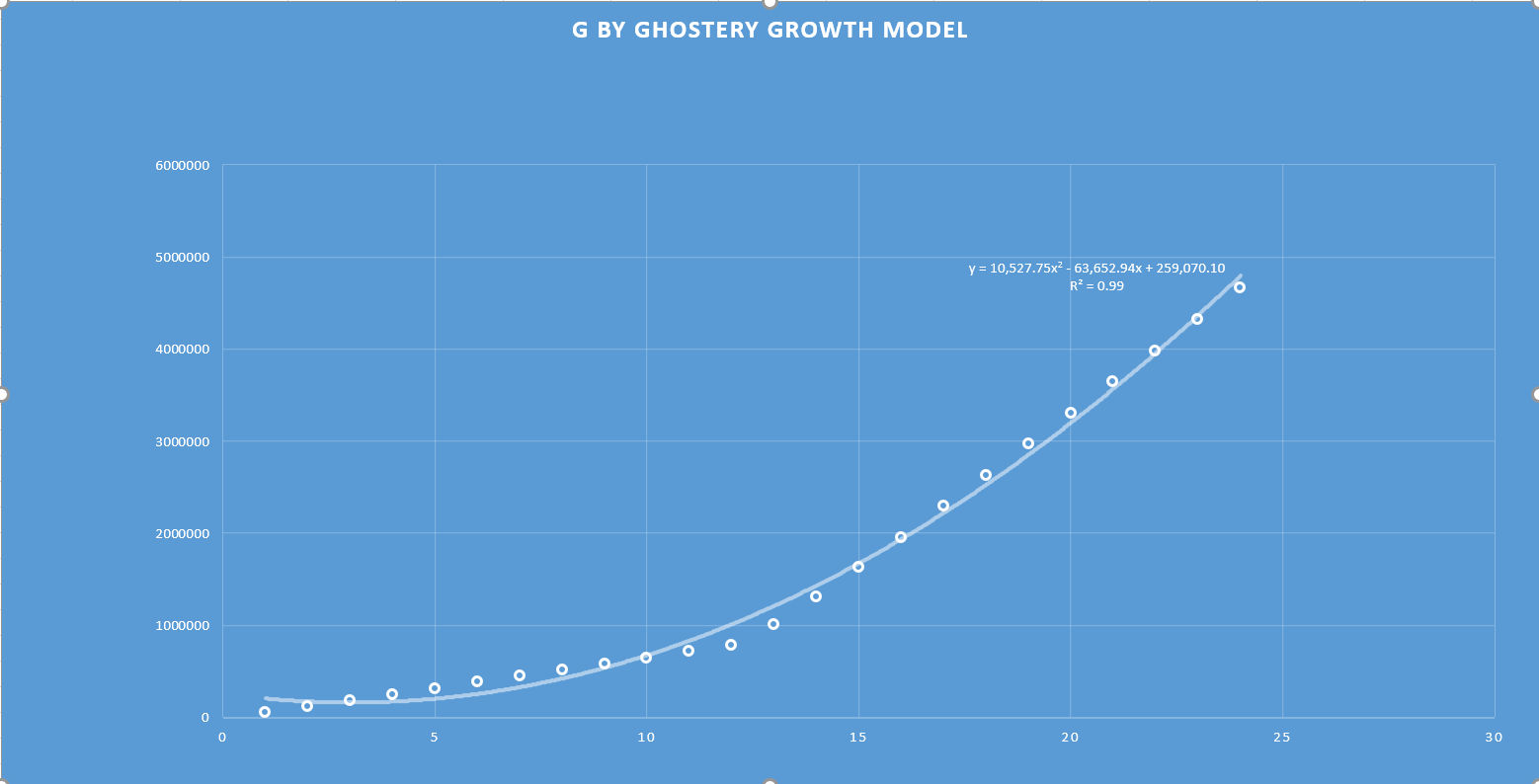
Therefore, we get 55,000 customers in the first month. Therefore, we end up spending $55,000 out of our $1,000,000 limit for each of the first 12 months.



For the next 12 months, our model will change slightly. We are assuming that second year will be associated with higher budget allocation towards taking this product to the next level.

* **New Launch Affect** stays 0 for these 12 months as well
* **Social Media** continues to give us another 2,000 every month
* **Partnerships** give us 500 more customers every month
* **Organic growth** – Word of mouth gives us 500 customers as well
* **K-factor***(growth factor)*-0.4 with the assumption that our product growth doubles on a monthly basis

As per our model, the **4.6 million** users by the end of 24 months.



Polynomial order 2 quite accurately fits all the points and would provide a good estimate of growth for the future also if the growth is sustainable.

1. What other data would you have wanted for your model that you couldn’t find?

-Location

-Mac Users

-Consumer/Business facing product

-Applications used

-Market Segmentation

-Feature of the new product

-Information for Competitor Analysis

-Go-to-Market strategy for previous launch

-Turnaround time : Time taken for our app to analyze website once URL is entered(Performance metric)

1. What difficulties did you come across in searching for data and creating your model, and how did you overcome them?

-Selecting curve for growth model – Fit best (Polynomial vs Linear) line for growth

Calculated values using equations and tried to minimize residual errors.

Analyzed R-squared values to find best fit.

-Finding Factors affecting success of product

Made assumptions on specialized costs and on-going market rates and metrics

1. What other factors are important to consider when considering releasing a new product? What are alternatives to accessing a more mainstream audience?

Our growth model has focused on the **marketing and budgeting** perspective but there are a lot of key factors to consider when releasing a new product

* Know your stakeholders
* Time of launch
* Pricing the Product ‘just right’
* Competitors in the market
* Focus on User Feedback

Our product is not the most ‘generic’ product. It is not something that would get used automatically.

One way to get addressed by the mainstream audience would be to make a **product which is easily noticeable and needed**.

The product we are to launch should **address customer’s needs**. For this, detailed market research should be done before investing heavily in the project.

Another important angle is to create something entirely new such as *flying cars.* Something like this would allow one to create an industry and disrupt the current automobile Industry.

Even when the product created does not bring the traction expected, there should be provision to ‘*pivot’* in order to create something more **meaningful**  and **relevant**.

Offering incentives, promotions is also a way of reaching wider audience. For eg, offering free access to premium version for, say 6 months, or keeping the product free for the first few months can also help to draw attention.

The most important and implicit way of getting mainstream audience is to **market the product like a mainstream product**. This can be done via social media ads or ***forming partnerships with mainstream companies such as Google for chrome and inserting this as a pre-loaded feature. The same approach can be forged with Apple and our product can be added as a pre-loaded feature with Safari.***

1. How would you measure the success of *G by Ghostery*? What data would you need to monitor these KPIs?

I would measure the success of this product based on Revenue, Total User Base and Active User Base for the first few months.

To measure the success of this product, I would follow and track the following metrics:

-Cost per Acquisition/Cost per Installation

-Retention cost per user(if any)

-No of new installations – Based on OS, Demographic, Age, Income

-No of new uninstallations - Based on OS, Demographic, Age, Income

-Cost per Click

-No of Applications blocked

-Average run time per user

-No of New Pause Blockings

-No of glitches after feature launch(Tickets)

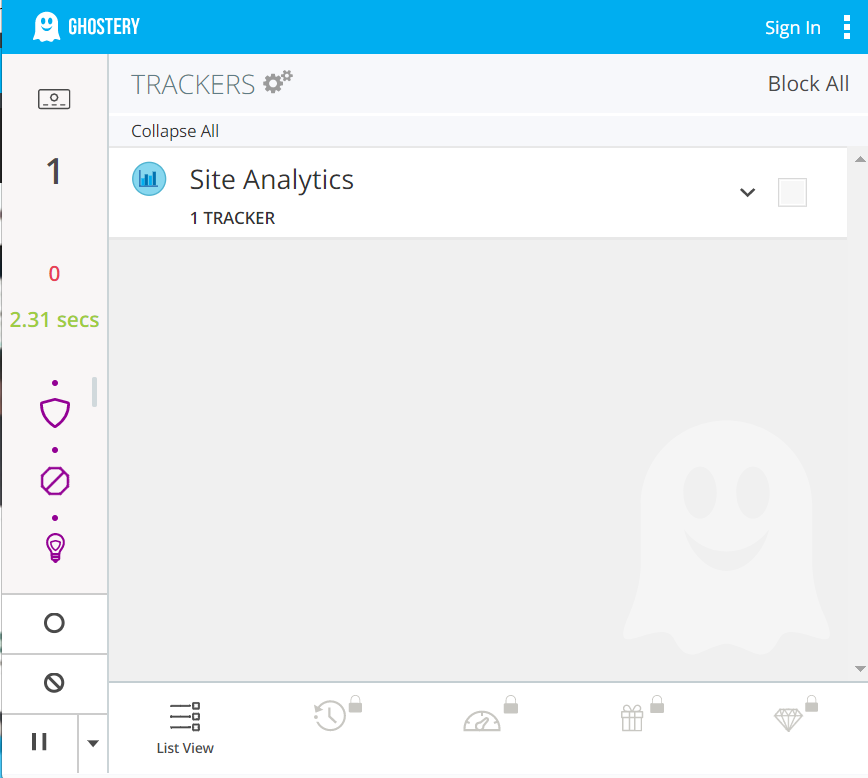
1. Assess the current Ghostery product and provide a couple of suggestions (features, design elements) to make the product more appealing to a mass audience.

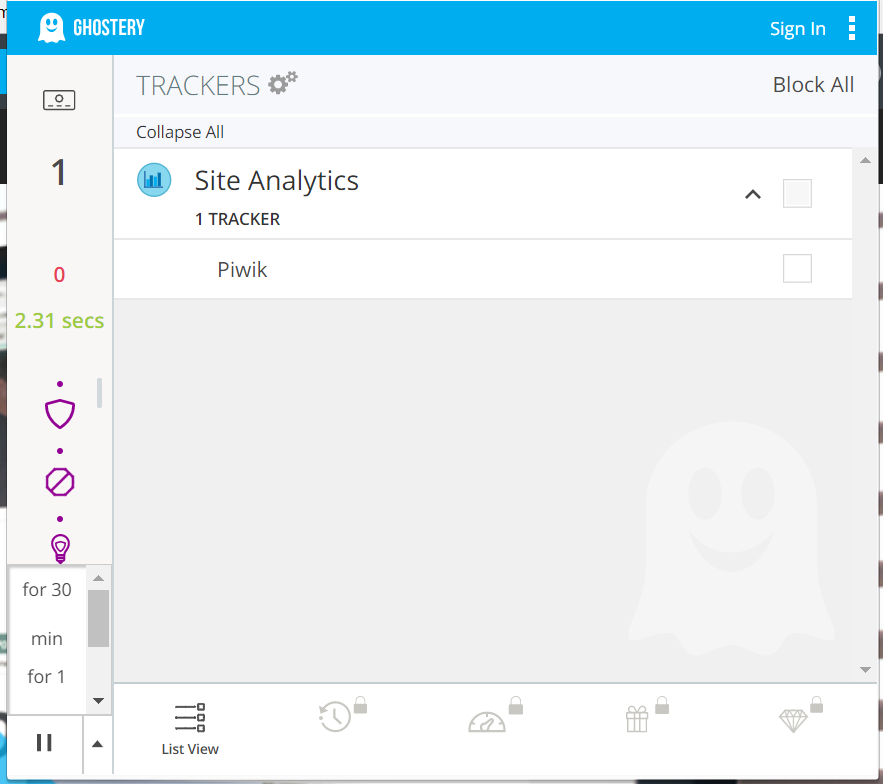
As of now, upon first seeing this dashboard, I feel a lot of things can be improved.

-This page has too many things going on for an extension

-Just one button is enough to start/stop blocking

-For in-depth analysis, user should be taken to website instead of showing **detailed** analytical data on a small window





The website and the first page that open were very clean and clutter-free. I saw scope of improvement in the above dashboard which is ‘*more advanced’*.

This dashboard looks like there is too much going on:

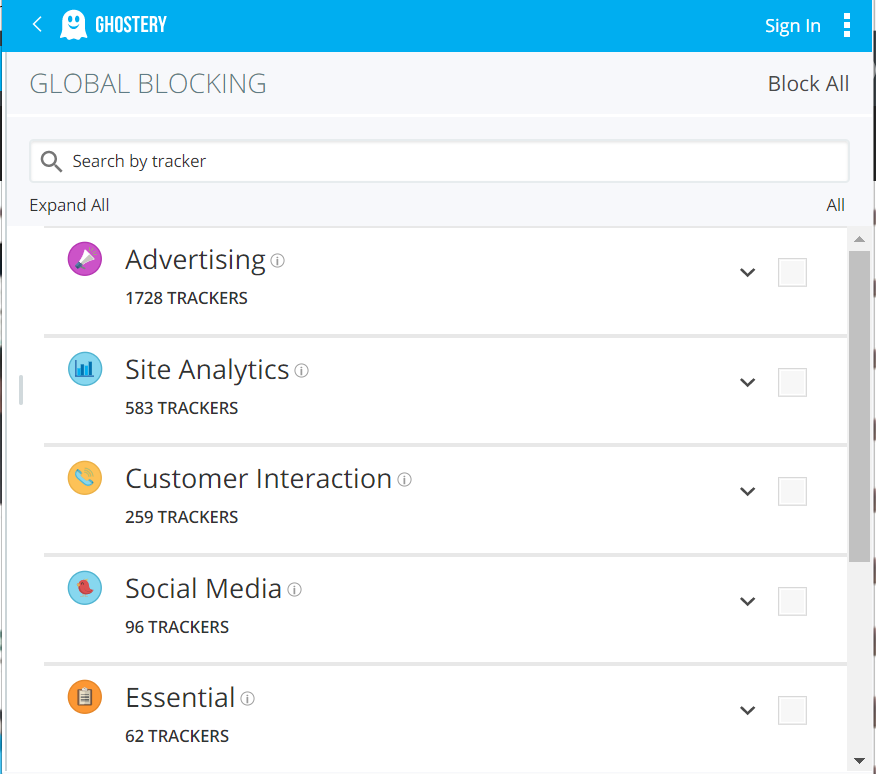
Suggestions and solutions:

-2.31 seconds doesn’t help a novice understand that it is the disabling time.

-A lot of icons on the left. Someone without any knowledge of Anti-Tracking and Anti-Browsing won’t understand what your enhanced anti tracking anti blocking features are.

-A better solution would be to have the Smart Blocking button on the top which changes color – say from Blue to Yellow when you press it and it’s activated and blue when it’s deactivated again. You can also have a slider button on top to switch it on/off.

-The drop down in the lower-left corner is not visually appealing. It extends upwards and blocks the view of the bulb. You could add these features to the already existing dropdown in the upper right corner.



This slide is not as important for a general user that we’re trying to capture. Someone who has no information about this Industry won’t really make much sense of it and may find it intimidating and off-putting.

In my opinion. There should be just one window which would consist of **easy on-off** button and a **graph** clearly depicting tracker usage. Everything else should be on a webpage which can be redirected by clicking an icon.

Something like this but with Ghostery’s logo would look amazing!

<https://www.w3schools.com/js/tryit.asp?filename=tryjs_lightbulb>

1. If you had more time, what else would you have done?

-Looked at the product from different angles

-Used the product for a week myself

-Market Testing based on features

-A/B testing by providing one button in place of 3-4 buttons as there are now

-Taken fewer assumptions and found accurate market costs

-Run statistical models on using pandas and sklearn on tangible data

Bring your product analysis plan to life with whatever assets, documents, and visuals that you see fit. Your work will be evaluated based on your thought process, creativity, thoroughness, and presentation. Be sure to clearly outline all assumptions.

# **Additional Data**

Figure 1: Ghostery monthly users by technical skill and device

Figure 2: Average monthly Ghostery users by year since launch

# **Scenario 2:**

# **Overview**

Imagine you are the Product Analyst at Ghostery and while you were on vacation, there was an influx of support tickets. The Product Manager wasn’t sure how to figure out what was going on, so he left you with the data to query when you returned to the office.

He provided you with 10 days of user telemetry data. Each observation in this dataset reflects one action by a user, and includes details on date, where they were, which operating system they were using and the version of the extension.

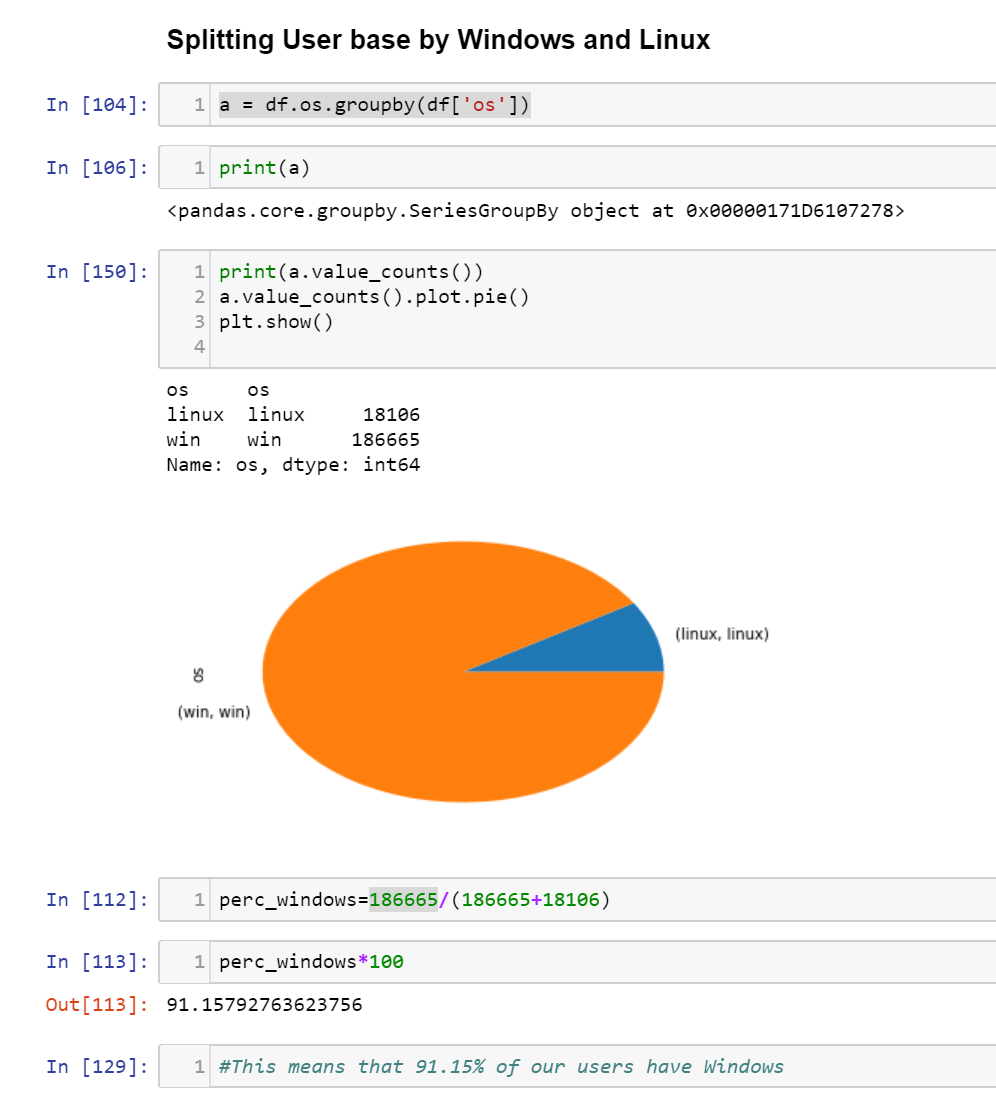
# **Deliverables**

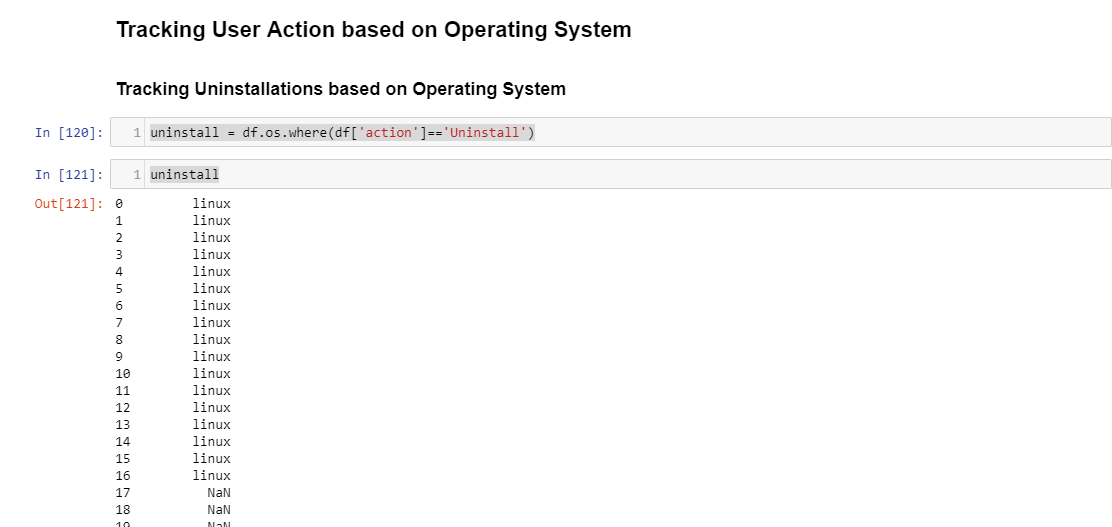
Provide an analysis of what you think happened while you were on vacation. Included in your analysis, please answer the following questions:

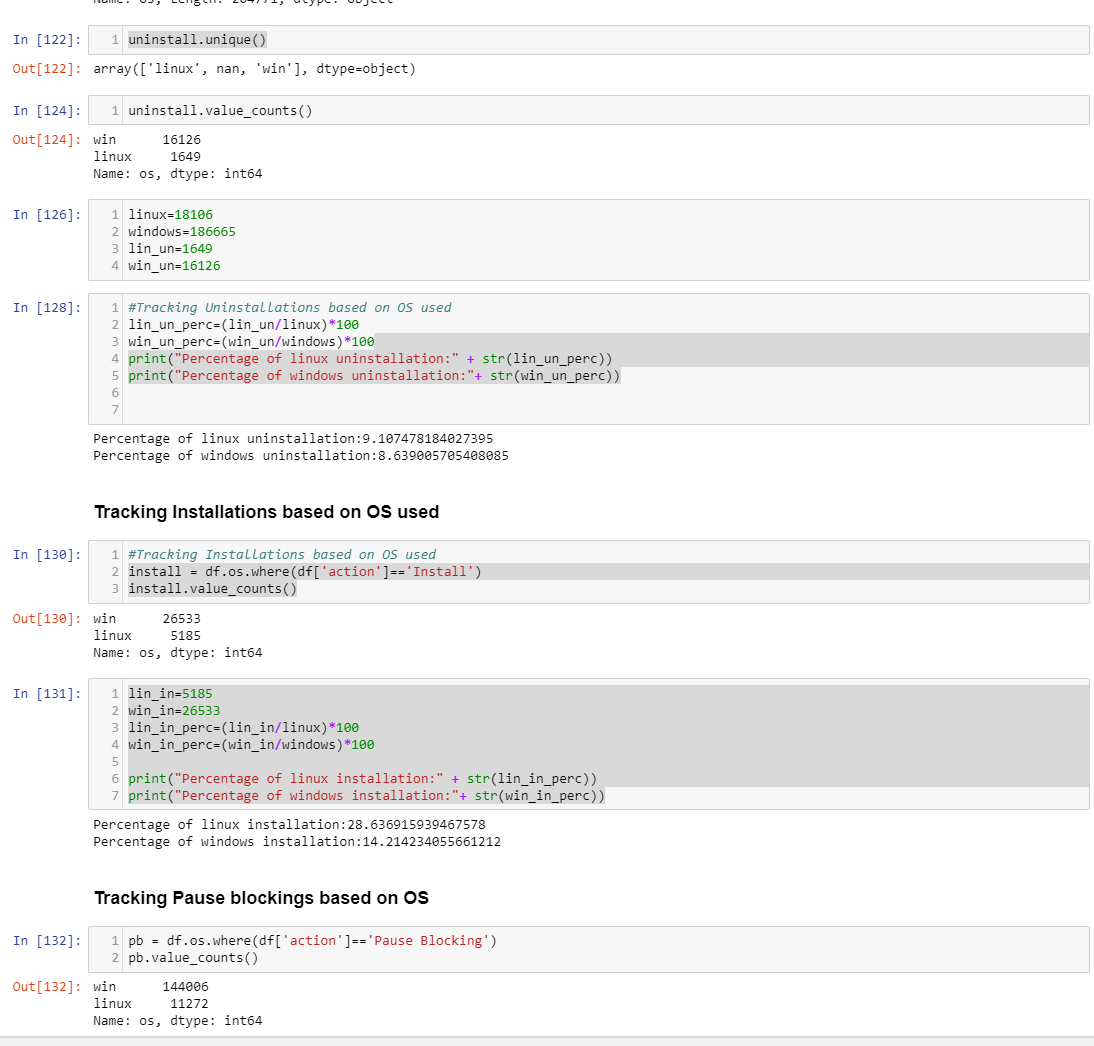
First of all, I had to set column names in excel file for which I used my own nomenclature. This didn’t work. I tried every trick. All columns were being merged as one and couldn’t be read by Pandas.

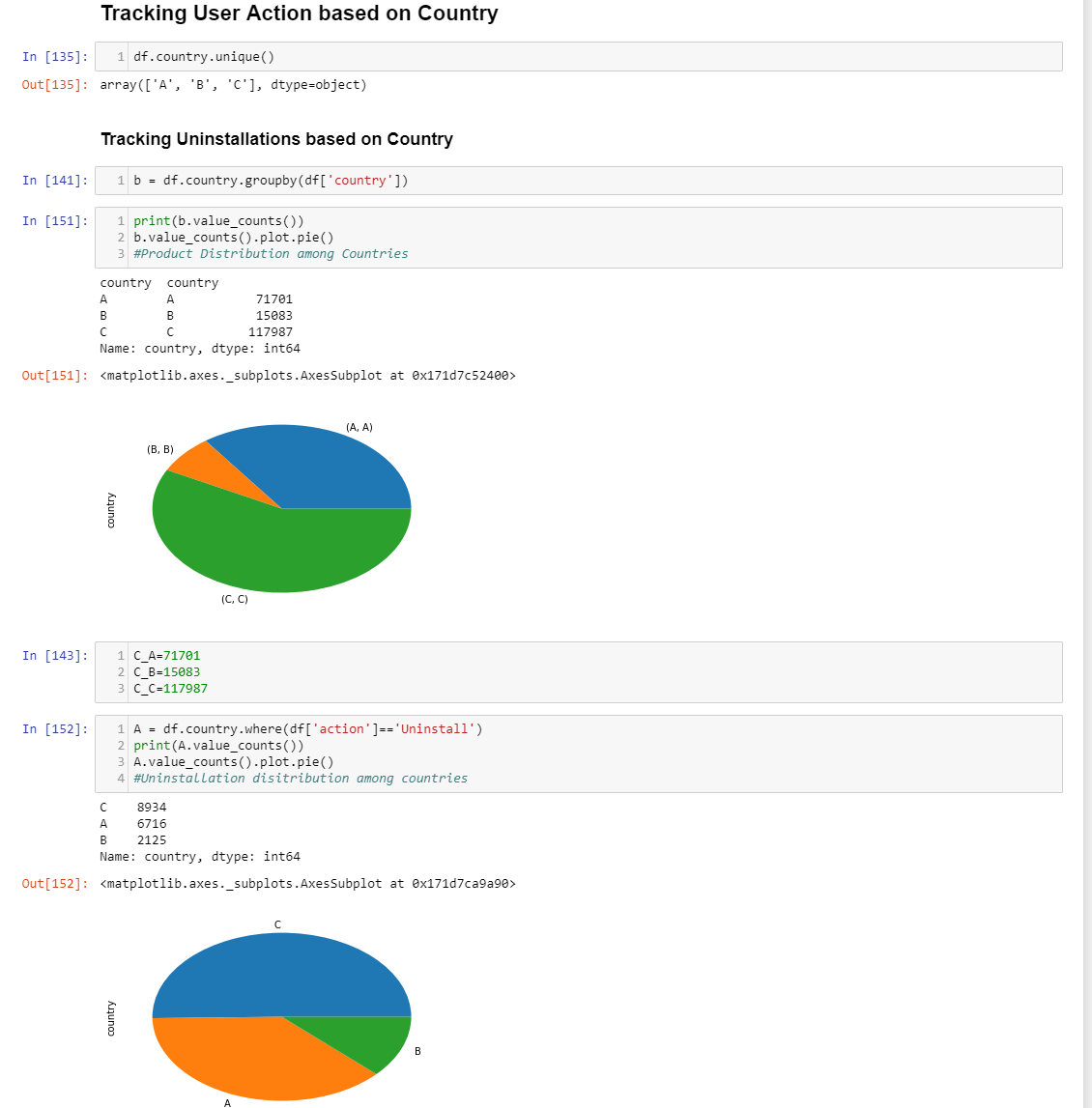
This was a huge problem. I found a way around this. I created separate data sets for each column and merged all of them into one dataframe.

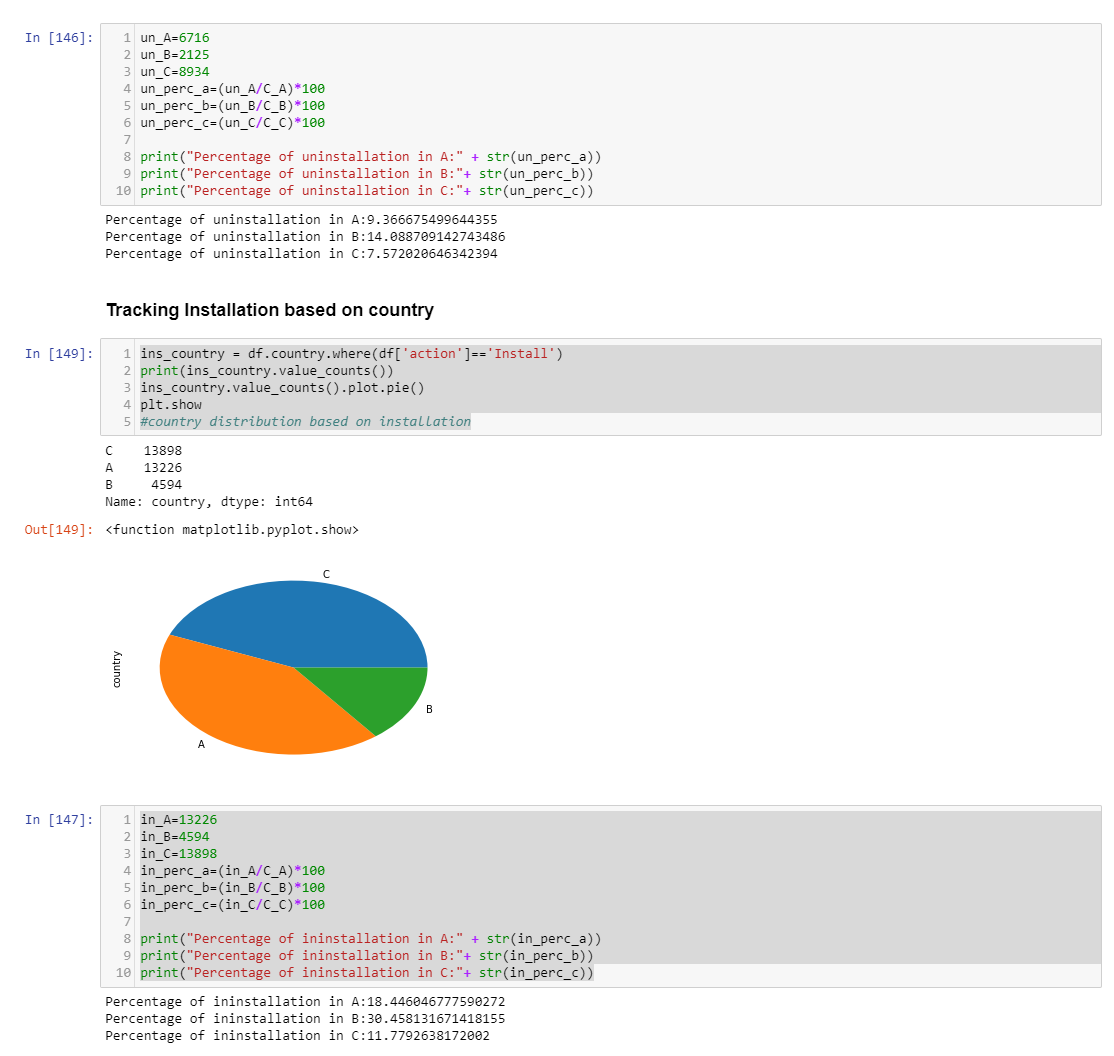
**Next, I broke down our users based on OS**

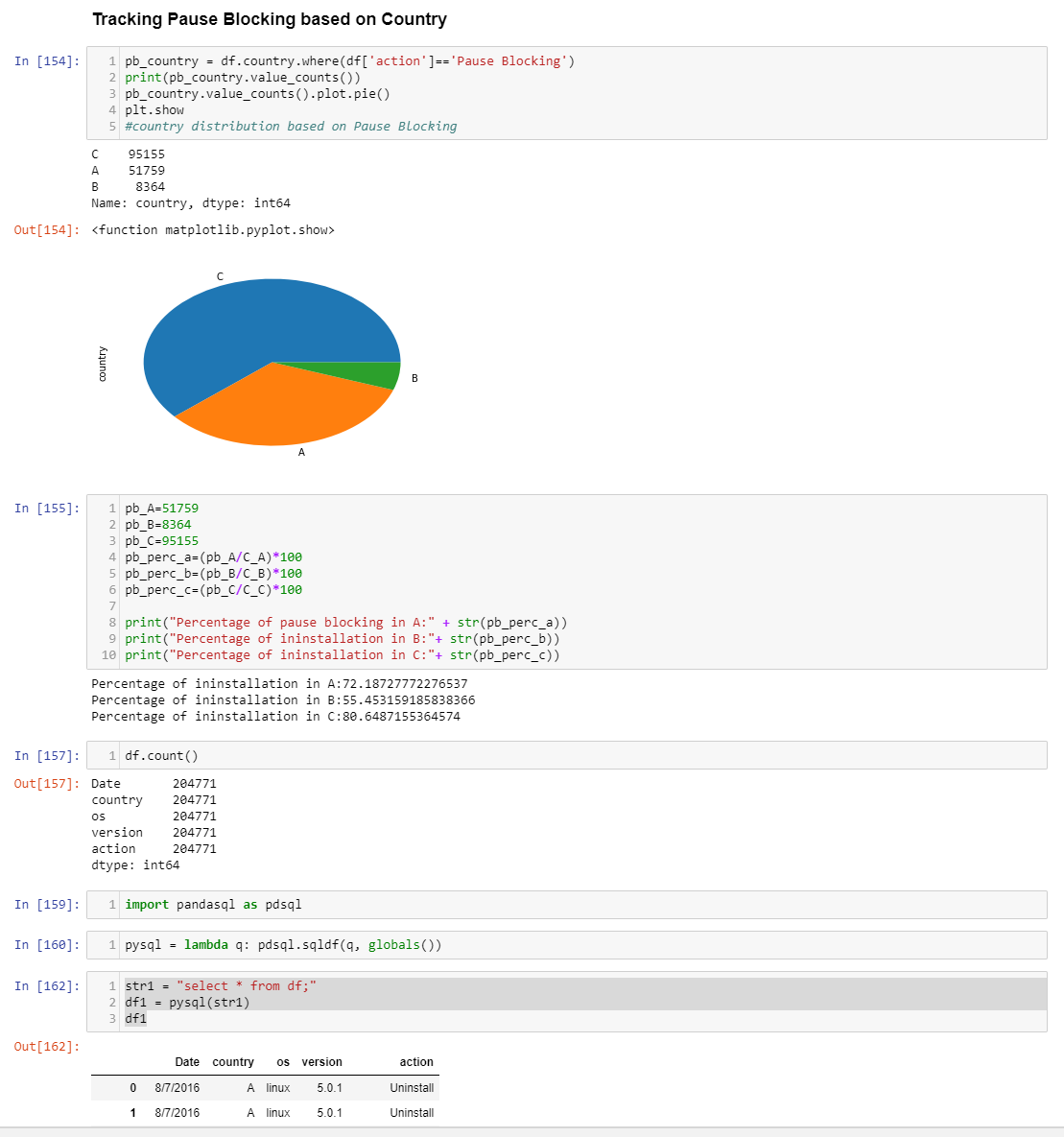






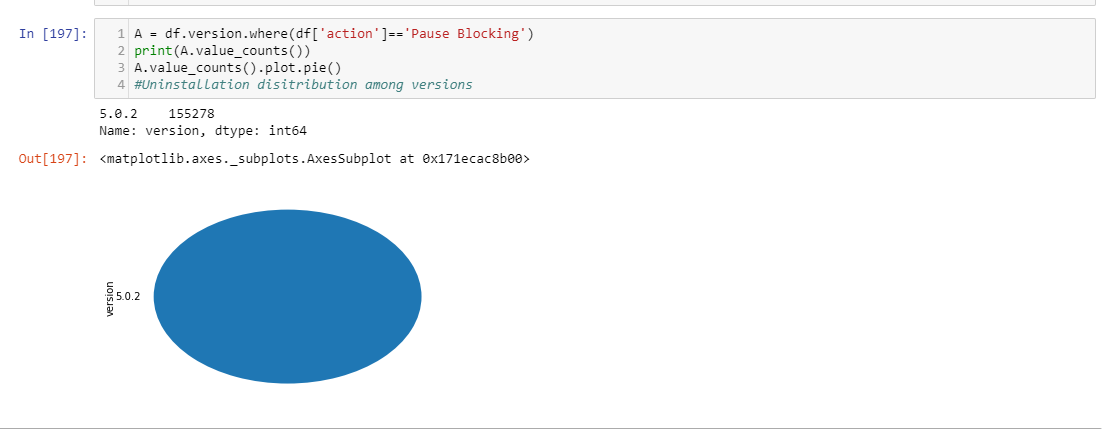






**Then, I finally used version as a parameter to gauge the problem**





1. What is going on? Why may this be happening?

I followed a step-by-step analysis tracking every attribute. First, I checked os, then country and finally version.

While country was evenly spread and windows to linux uninstallations even, I noticed that most of the installations were made with version 5.0.2 on windows and 100% of the pause blockings was on 5.0.2 version of Windows.

This may be happening because the new version release 5.0.2 must be having some technical bug. There may be lag, or it may not be performing task as well as it should have been and its overall performance must be worse than 5.0.1

1. Is the activity you are seeing related to a specific parameter?

Yes, as per my analysis, it is due to the new version release i.e. 5.0.2 and most of the installations and all of the pause blockings are due to problems with this version.

1. What interesting ways can you visualize the data?

I have visualized the data as pie chart and histogram(below) to see how uninstallations and pause blockings are spread across version and OS.

1. What other data would you have wanted for your investigation? Can you think about a way you could get this data?

-I would have wanted comment or reason for the uninstallations/Pause blocking.

-Also, I would have wanted unique IDs for these uninstallations because a lot of this data could be redundant and it could be possible that it was entered erroneously (for eg 10 tickets were raised per installation). Unique ticket id linked to user id would help solve that.

-Some information about user(such as age, user rating, location, etc.) would have helped in root-cause analysis.

1. What other questions do you have?

-Reason for Uninstallation/Pause blocking

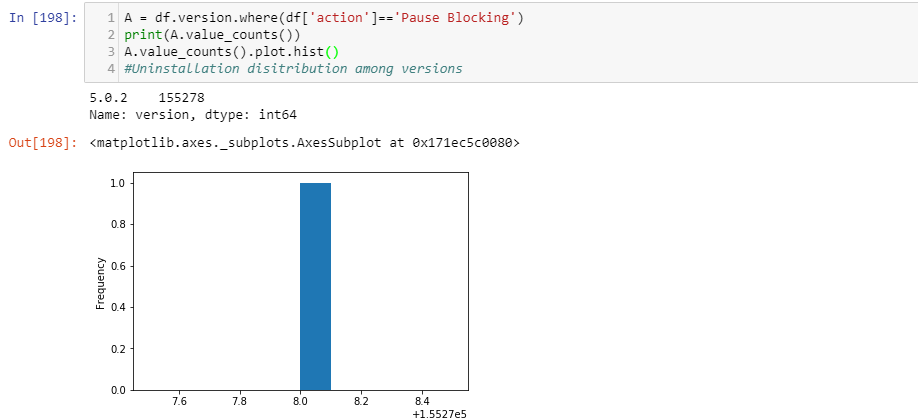
-Are there any Mac users that we should evaluate?

-Is our entire customer base comprised only of Windows and Linux users?

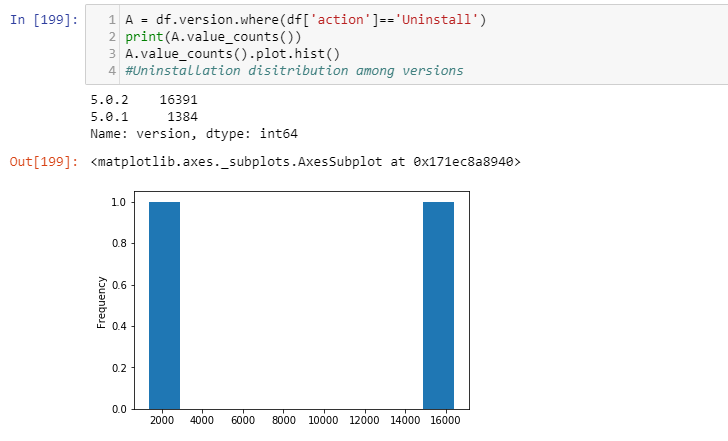
-How long have the users who have uninstalled been using our product?

-Did they uninstall previous version?

-Have there been similar problems with previous versions?



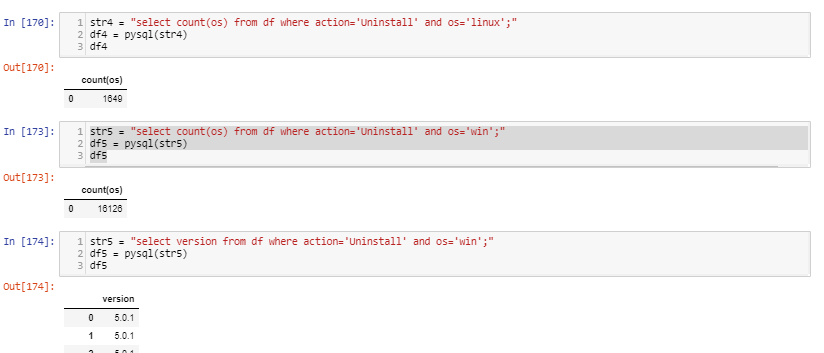
Histogram depicts all Pause Blockings due to version 5.0.2 and on Windows



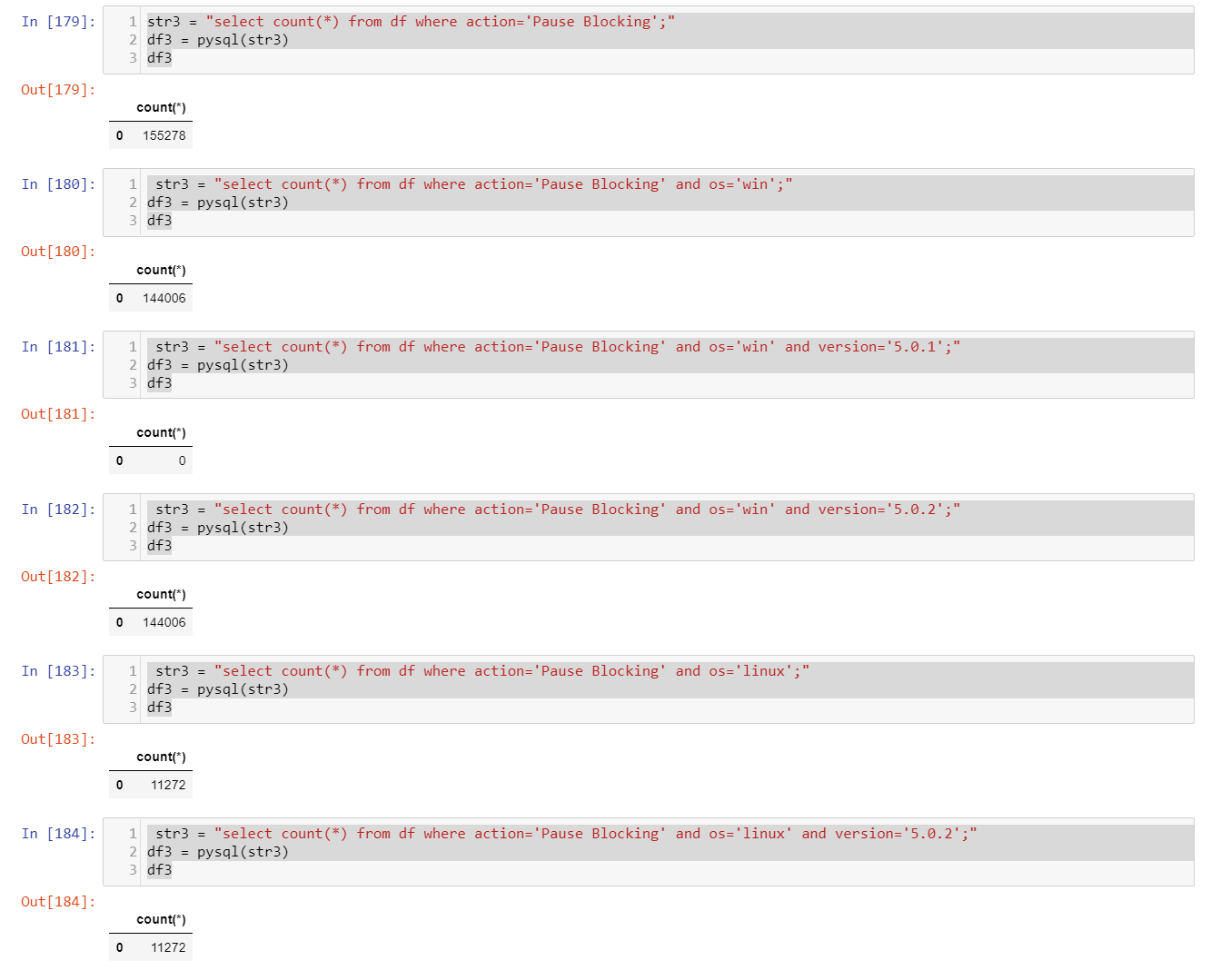
This shows split in uninstallations based on versions.

Here are a few SQL queries I ran with pandas to verify my claims:

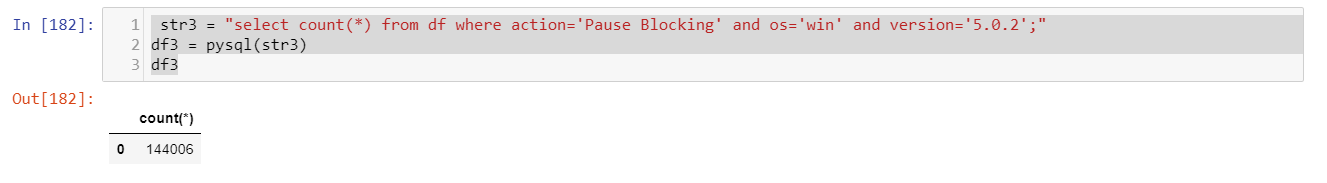
**Uninstallations based on OS**



**Pause Blocking based on OS**



**Digging deeper to find combination of OS and Version causing major problem**



Feel free to use whichever coding language(s) you are most comfortable with. Be sure to include your code (clearly commented), effective graphics to support your findings and a summary of your observations/analysis.

# **Additional Data**

Please see attached csv file ‘scenario2-data.csv’. The order of the parameters are as follows: date, country, operating\_system,version, user\_action.