Dear Game Developer's Business Team,

I am presenting you my report. As you know from my resume, I have never used Power BI before but since you had recommended to use it as a part of the present task I decided to try. I spend last two weeks on learning it a bit and the result is enclosed 😊.

//Unfortunately, I cannot share the report and dashboard with you trough Power BI services as we are not in the same domain. That is why I am sharing the report with description in .pdf file and the .pbix file with the report.

REPORT

The data represent game sale event lasting in December and January in **two** following **seasons**, i.e. **December 2016-January 2017** (season #1, s#1) and **December 2017 – January 2018** (season #2, s#2) (see Fig.1.1). **S#1 was much better** in terms of sales **than s#2**. It is true for all regions analyzed excluding Eastern Europe (EE); in Eastern Europe s#2 was better than s#1 (Figs. S1 a)-b)). For this increase in EE region responsible is Poland where the sale increased significantly in s#2 compared with s#1. **Total Net Sale (USD)** (whole data on s#1 and s#2 summed) was **1599434.64 $,** wherein **s#1** contributed to in **63%** (1009595.87 $) and **s#2** in **37%** (589838.77 $). The period in both seasons that was **the hottest** in terms of net **sales** was timed between **December 21st and January 3rd**, so the Christmas – New Year's beginning.

Generally, in two seasons, the **top platform** was **Windows**; Mac was the second and Linux the last (Fig. 3.1 – 3.2) The popularity pattern varied a bit depending on the region, e.g.: in Africa and Latin America (Honduras) the Mac dominated. Nevertheless, there were only few units sold totally in these Mac domination regions.

In two analyzed seasons, **top sales** region was **North America**, wherein the top country was **USA**; the **second top** region was **Western Europe** with **Germany** as top sales country (Fig. 3.2-3.3). There is no such a region in one season that is not present in the second. This implies the geographical pool of targets are similar in both sale events.

The amount of **Returns** (units) were averagely **10.15%** (s#1) and **13.68%** (s#2) of Gross Units Sold (whereas for two seasons together it was **11.6%**). **Returns are independent on Platform**; for Windows, Mac and Linux there are respectively 11.76%, 9.87%, 5.05% Returns of Gross Units Sold. These differences in % values are assumed here as not significant since the more units sold, the more returns may happen, simply, and Windows is clearly predominant platform in the present data set. The contribution of Returns in Gross sale can be assumed as quite constant and in the future predictions should be ranged between 10-15% for a season (Fig. 2.2). On the basis of provided data, **the Returns are not significant** in terms of **differencing the seasons** or **explaining the net sales progress** (Fig. 2.1 a)-b), Fig. 2.2). However, note that the higher contribution of Returns in Gross Units Sold weakens the already weaker season s#2. Here the analysis of the Returns' reasons could bring more information however in the set there is no data on it.

The **discount** applied during the sales period **varied over time** and **region** and ranged between **0.00% and 76.74%** (e.g.: Figs. 1.3a)-b), Fig. 5.1, Fig. S2.2)**.** The **average** discount percentage was **17.44 %** (this is the average over the whole data set). Figs. 1.3 a)-b) show that **50%** of the data had the discount lower than **9.74%** and **11.62%** applied, respectively for s#1 and s#2. Generally, for two seasons together, **56%** of the data **had the average discount lower than the threshold value of the preferable range** (i.e. lower than 32% (Fig. 5.2), see further in the text). Visualizations supported with statistics say that the **discount is the key parameter** having the

impact on sales effect. However, it does not mean that gross sale (and net sale) and discount value are linearly increasingly correlated. **There are ranges** wherein the **sale is the highest** and some wherein the **sale is the lowest or lower than average**. **Key Influencers** (KI, Fig. 4.1) analysis shows that the **most effective sale** is when the **discount** ranges between **32-45.27%** and **the less effective** when **discount** is **more than 45.27%** (and also between **0 and 20%**). The range 32-45.27% is a segment where the average Net Units Sold (27.9) is higher than the total average (12.91, the average over the whole data set); this refers to 34.4% of data. Moving on to less preferable segments, the segment of the discount value higher than 45.27% has the average Net Units Sold of 2 and refers to 6.2% of the data, whereas the segment with discount value less or equal 32% has the average Net Units Sold of 5.36 and refers to 59.3% of the data. However, I suggest not to take this information outright. When we compare the hottest sale periods in two seasons (Fig. 5.1, Table 5.1-5.2), we can notice that despite higher discount values in s#2 in that period, s#1 still wins. It can be a reason of the product type itself and the characteristics of sale event but this info is not provided for in the data set. The **discount** of about **30-35%** seems to be **perfect** and applying the discount during the **longer holiday time** (few days off work and school) **seems to be the best idea**. Note that the **discount event** within the sale period should **last at least few days**. Applying a high or even the most preferable discount for one day only does not make a progress (Fig. 1.4, Fig. S1 a)-b))

If nothing changes and next season takes place with identical conditions it will be worse than s#2 in terms of sale progress.

If I were in charge of advising the plan for future sales of this kind, on the basis of the data I was equipped with, I would recommend to think about adjusting the discount values and periods of discounts. 50% of the data has the discount values placed in one of the less preferable range (0-20%); the average discount for the whole data set is 17.44%. The discount about 30-35% during Christmas – New Years beginning time seems to be an accurate choice. I would keep it. However, I would think about the second discount event as an up-sell somewhen at the end of January when most of the countries have different kinds of winter holidays. Note the discount event should last at least 4 days, not shorter since the discount episodes (1-2 days) are not effective (can be easily missed). What is more, I would think about the marketing activities and compare how they looked like before s#1 and s#2. I would also have a deeper look the market that made much higher progress in s#2 than s#1, i.e. EE and Poland within.

Summary

- There were two sales seasons (s#1 and s#2). The first was much better than the second; the contribution of s#2 to total progress was 63% and the contribution of s#2 equaled 37%.
- The hottest sale period in both seasons timed between December 21st and January 3rd.
- The top region was North America with USA as the top country and the second top region was Western Europe with Germany as top country. This referred to both seasons.
- The top Platform for both seasons was Windows. It was generally the worldwide pattern.
- The amount of Returns (units) were averagely 10.15% (s#1) and 13.68% (s#2) of Gross Units Sold (for two seasons together it equaled 11.6%). On the basis of provided data set it can be claimed that the Returns are not significant in terms of differencing the seasons or explaining the net sales progress.
- The discount applied during the sales period varied over time and region and ranged between 0.00% and 76.74%. The statistics showed that discount value had a significant impact on sale efficiency. The analysis demonstrated that the most effective sale was when

the discount ranged between 32-45.27% and the less effective when discount was more than 45.27% (and also between 0 and 20%).
- If nothing changes and next season takes place with identical conditions it will be worse than s#2 in terms of sale progress.

Answer to bonus question:

On the basis of the present data set I can only assume that the numbers refer to a cyber product (online game) which was probably released on 2016 and upgraded next year.

For more detailed analysis, some additional data would be needed.

If this were a real-world case I would ask for:

- data on the product itself (Is it a cyber or physical product? Is the s#2 product a new version (second release) of s#1 product or is it just an up-sell of exactly the same game? What is the name of the game? Is it winter-time/Christmas game for kids or rather battle/racing game, theoretically for adults (in my opinion the latter is the true)? The "Product ID" param indicates all the records refer to one product and there are probably no different versions within and between the seasons. If that is true it could be one of the reasons for s#2 being worse than s#1.
- data on the seasons (Was 2016/2017 the first sale season for this game? Will there be another sale event for the product? Was the game available to buy before December 2016?)
- data on the marketing activities realized before s#1 and before s#2
- data on the reasons of returns (whether there is a base with customers complains/reasons of returns)
- data on customers (users) (age, male)
- data on orders (For example a purchase on a specific day in a specific country shows 5 units sold – How many customers did the purchase on such a day in such a region (1 or 5, or maybe 3 customers)?)
- data on data construction (What does "Package" param mean and why is it always "-1"; is it because of the product being e-product? What does the "Type" param, which is always "direct", refer to?)

Answer to bonus bonus questions:

I think the data represent the sale of an online game. The data may originate from a big platform, for instance Steam, prospering worldwide and offered for three operating systems: Windows, Mac, Linux. I exclude Origin because it does not serve Linux platform.

The sale events represented in data are probably annual winter sales, timed around Christmas and New Year's beginning time.