

LSE DA301 Presentation

By Kevin Wijaya Oey

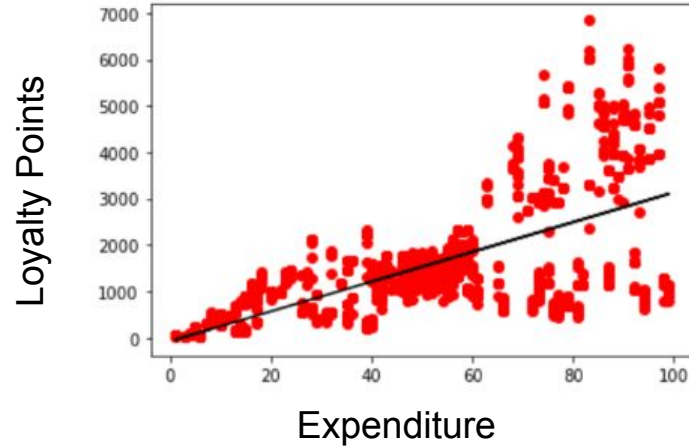
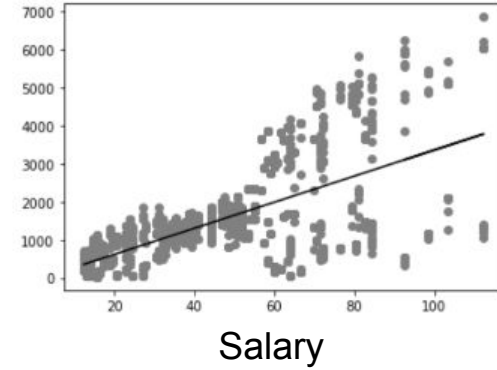
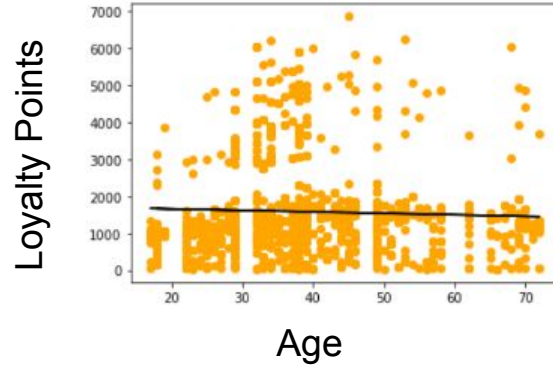
Objectives

- **Product Categorisation**
 - find popular products
 - remedy customer complaints from bad products.
- **Optimal sales and incentivisation strategies**
 - identify strong predictors of loyalty points
 - Target lucrative market segments (for long / short run).
- **Evaluating data reliability and correlation** to gather more accurate data and create better predictions.

Approach

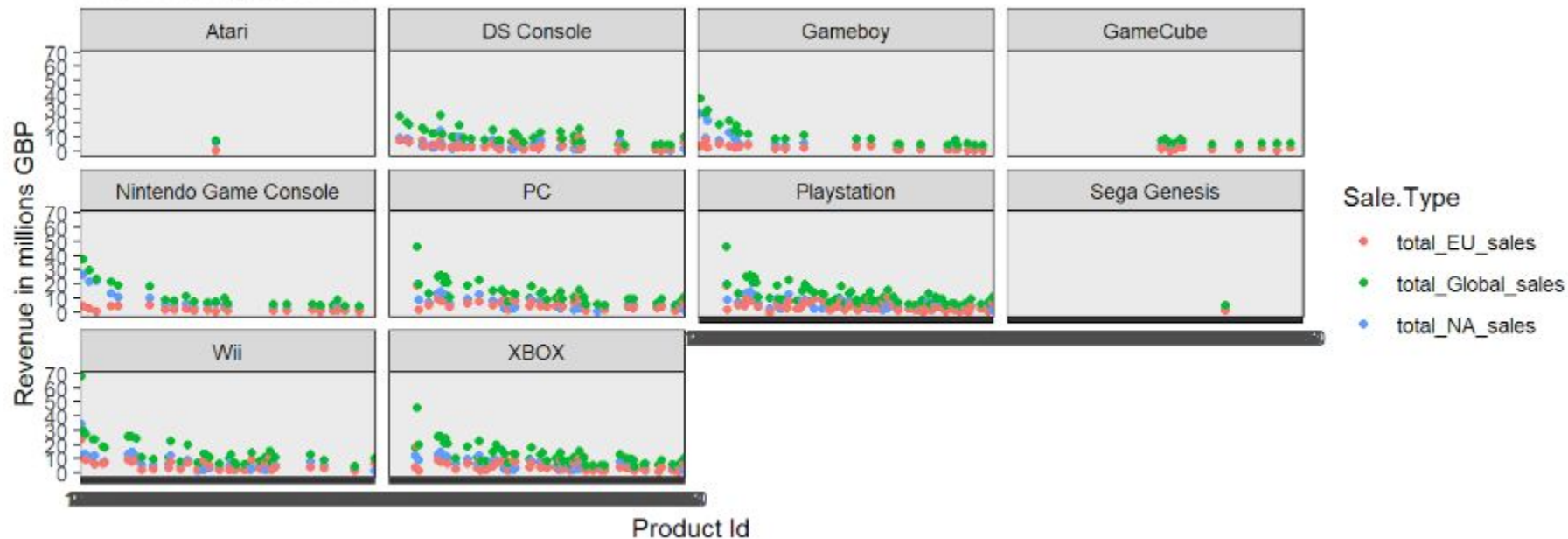
- Linear Regression - to find relationship between (age,expenditure_rank,salary) against loyalty points
- k-means clustering to determine critical customer segments to maximise revenue.
- Textual Analysis:
 - Determined popular words
 - Checked sentiment associated with reviews and words
- Created complex charts to see how products influence sales.
- Determined data reliability (kurtosis,skewness, shapiro-wilk tests and qqplots)
- Created a multiple regression model to predict the influence of North American and European sales towards Global Sales.

Results - Linear Regression



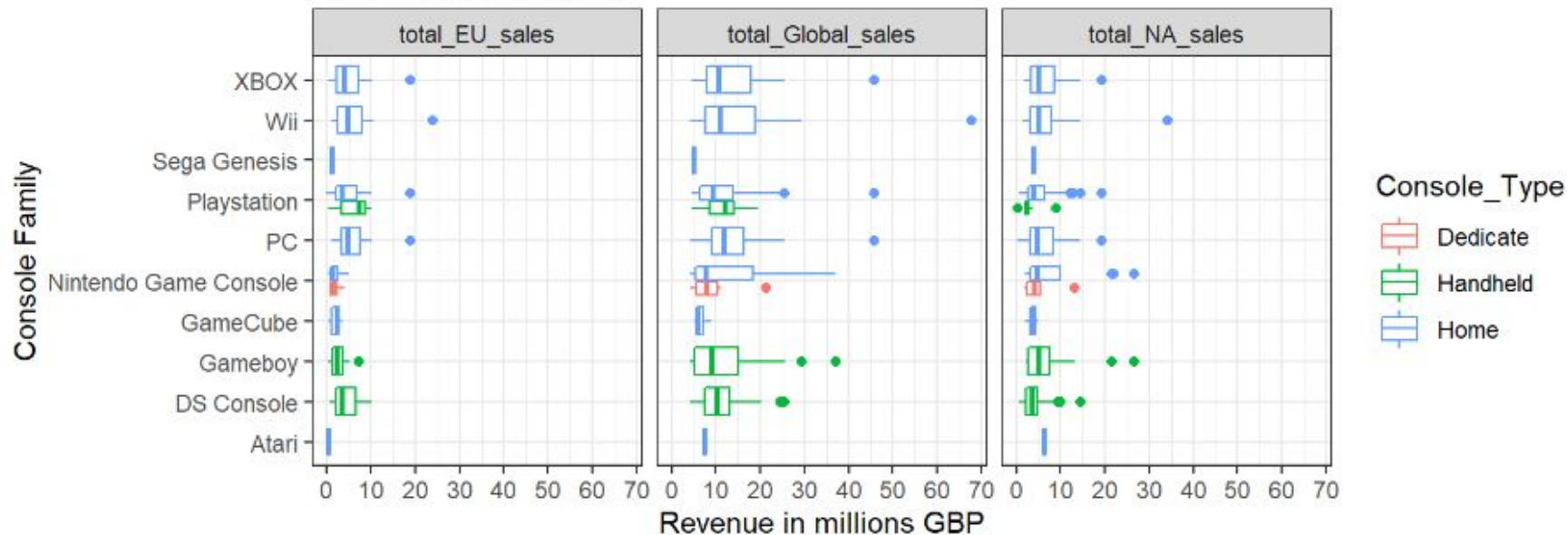
Breakdown of Regional Sales by Console Family

Source: turtle_sales.csv

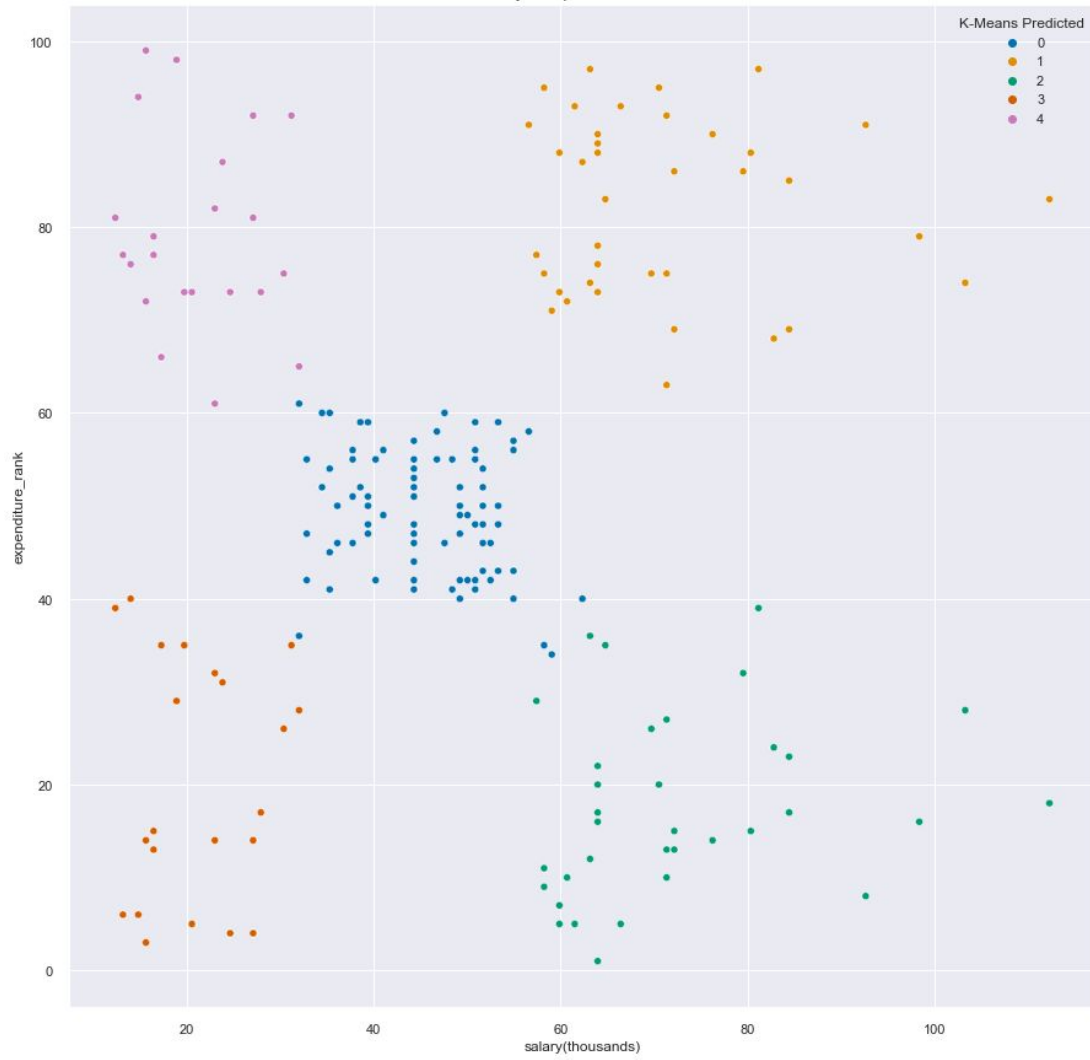


Boxplot of Sales by Region segmented Console Family & Console Type

Source: turtle_sales.csv



Salary vs Expenditure Clusters



using little go use enough feel rule people set many new long
product strategy keep ball thing puzzle far end toy play try say probably monster sticker children
question two find box like player point kit give card look mini hard playing lot
could three different every work loved word kid help well enjoy year easy
quest still pretty see year old color day put buy didnt would version everything used definitely board game part come without want hour seem adult adventure
back need though expansion getting review way plastic old color day put buy didnt would version everything used definitely board game part come without want hour seem adult adventure
much family think fun price add together nice however first open quite around bit since egg building made size
acquire another anger friend best overall good looking hand already however first open quite around bit since egg building made size
great love son dont time really book
gift piece love son dont time really book
able cardboard ball whacks instruction making original something great game miniature doesnt animal

Top 20 reviews and summaries

The top 20 positive reviews and summaries show that customers are satisfied with their purchases. The negative reviews show that customers are mostly dissatisfied by:

- incomplete product packaging
- unclear instructions
- products not working properly
- incomplete kits
- unclear messaging (eg games need patience)
- copied games or quality issues

Data reliability : Multiple tests confirm that the data is not reliable.

Predicting Global Sales Based on EU Sales

total_NA_sales	total_EU_sales	predicted_global_Sales	Actual_Global_sales	Prediction_accuracy
34.02	23.80	67.510029	67.85	-0.5010633
3.93	1.56	7.176902	6.04	18.8228762
2.73	0.65	4.744629	4.32	9.8293779
2.26	0.97	4.559482	3.53	29.1637944
22.08	0.52	27.168021	23.21	17.0530849

Average Percentage Error: 15.1%, still needs some fine tuning



RECOMMENDATIONS

PRODUCT & SALES STRATEGY

- Contact disappointed customers and remedy their concerns.
- Investigate the games that have product issues and triangulate pain points - is it a supplier issue, or are we mispackaging things?
- Review unclear or misleading instructions and packaging.
- Continue to market games and books, especially, Playstation, Wii, XBOX, PC and Gameboy consoles to customers with high salary and high expenditure and moderate salary and moderate expenditure. This is especially pertinent as customers who spend more have higher loyalty points.
- Stop selling Gamecube, Atari and Sega Genesis consoles.

DATA ACCURACY

- Check the feedback systems as there are duplicate reviews.
- Consider procuring better datasets.
- Add more fields, such as regions, store names and vendor suppliers to add granularity (eg to investigate where the bad products come from).
- Provide data on date of customers' purchases so we can identify which customers return over the past 6 months.

OTHER AREAS OF EXPLORATION

- Test the linear regression model with other variables and evaluate its usefulness with a confusion matrix.
- Provide contextual analysis in addition to sentiment analysis for more accurate results.
- Apply sentiment analysis on data scraped from social media sites and competitors' sites.



Thank you! For questions please contact:
Mobile: +12345678
Email: kevin@turtlegames.com