# Artists deaths and their impact on artwork sales

**Nicolas Brunner** **Quentin Bouvet** **Sandra Marcadent**
quentin.bouvet@epfl.ch
nicolas.brunner@epfl.ch sandra.marcadent@epfl.ch

## Abstract

This project aims to study the effect of an artist's death on his artwork popularity, by using reviews count on Amazon as a metric of popularity. It makes use of the Amazon reviews data from 1996 to 2014 and of current Wikipedia online data. Those two sources were combined to make the link between deceased artists and their derived Amazon products and reviews. The resulting data was analyzed by computing custumer interest in the form of reviews count per day, for artists deceased between 2003 and 2013. Subjects time series were averaged and compared with controls time series, which indicated a non negligible increase in reviews count following the general death time. The analysis was restricted to musicians and actors, both being treated separately.

## 1 Credits

This project was performed as part of the course Applied Data Analysis (ADA) taught by Robert West at Ecole Polytechnique Federale de Lausanne (EPFL). Authors are master students Nicolas Brunner, Quentin Bouvet and Sandra Marcadent, from the Communication Systems, Computer Sciences and Bioengineering sections respectively.

## 2 Introduction

It is often said, ironically, that Van Gogh never sold a painting in his lifetime while he is one of the most famous painters in History. Is it possible that people feel more interested in artists artwork after their death? Does this kind of effect plaz also to modern artists and is it more marked when the news is recent?

In the web medias, different articles describe the phenomenon by which decease of some artists generated a huge increase of their artwork sales **??**. The main reason invoked is the increased mediatisation following the decease and the exttend of advertising invested at this moment by the producers.

However, this is about really specific icons and those articles dont provide a serious description of the analysis made and the data used. Therefore, in this project, we wished to make a robust analysis of the effect of artists decease on their artwork sales by the use of Amazon reviews count as a metric of sellings. This research was applied to artists with varying popularities, deceased from 2003 to 2013.

## 3 Methods

### 3.1 Collecting data

#### Amazon Review Dataset

For this project we used the Amazon review Dataset. We considered for our model the reviews linked to actors and musicians. So 4 categories stand out of all those from Amazon: *Movies and TV, CDs and Vinyl, Digital Music, Amazon Instant Video*. Furthermore we needed the list of all the products with all its metadatas. The same site provide also products metadatas, so we retrieved those in the same categories as the reviews.

#### Amazon Product API

The metadatas provided was not enough for our work since we wanted links between products and actors or musicians. This is why, the Amazon Product API was required.

#### Wikipedia

The list of actors and musicians dead between a period of time (1996-2014) was needed and the simplest way to obtain this information was to scrap it from the wikitext of Wikipedia. Wikipedia contains summaries of every year data, including a list of relatively famous dead artists. From this list we scraped data using regular expression to extract: the *name, birth date, death date, description, and 2 booleans indicating If a subject is an actor and/or a musician*. Those 2 booleans were computed from the description, which was in general pretty brief, and thus keyworks detection was enough to compute those results.

### 3.2 Matching data

For each category, the 3 datas composed of the list of artists dead between the required time period, the metada completed with name of artists linked to a product and the dataset of products reviews, were cleaned and formatted to build a global dataset, matching subjects and reviews. The list of control names was constructed from the metadata with completed information about artists names, filtering out the artists dead between 1996 and 2014. This list was subsampled randomly but all keeping a ratio 4:1 of controls to subjects in order to maximise the possibility to match correctly them correctly. After the matching, the controls group size was reduced to that of subjects and paired information was kept. For more details on the criterion for matching subjects and controls, see section 3.3.

### 3.3 Data analysis

For each subject, a time series of reviews count per day was computed within a time window surrounding the death day. Subjects with really low popularity at the time of death and therefore not susceptible to have a mediatized death, were filtered out. This was done by imposing a minimum of 100 review counts during this time window.

Each individual time series was corrected from the general Amazon reviews count tendency, which is the reviews count per day for the total products of Amazon, in the corresponding category.

As can be seen on the figure 1 the total review counts per day for all Amazon products increases considerably after 2012 for Video and Movies (blue) and a little increase is observed for CDs and Vinyls (orange), which illustrates the need for a normalisation of data to avoid this bias. This was computed by multiplying the time series by a factor defined as the total number of reviews the first day of 2014 over the total reviews count at the current time:

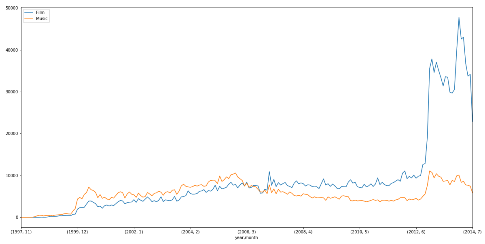


Figure 1: Tendency curve for general amazon product reviews count.

$$corrected(t) = \frac{tot\_counts(1/1/2004)}{tot\_counts(t)} \times count(t) \quad (1)$$

The time window for the analysis extends from 6 months before the death (-180 days) to 6 months after the death (+180 days) of each subject. Otherwise, each subject was matched to a control and the control reviews count per day was computed within the same 12 months time window, centered on the death date of the corresponding paired subject.

To ensure an unbiased comparison, the controls and subjects were matched based on their reviews count per month during the 4 months preceding the death of a candidate subject. Indeed, to build a pair, these simplified time series were vectorized and the distance, as a L2 norm, was minimized. This was performed to ensure that subjects and controls have comparable popularities but most of all that their time series vary in a similar way before the death.

At the end, each artist time series (subject or control) was normalized by its overall popularity defined as the individual total reviews count within the 12 months time window, in order to avoid bias due to variations in popularity.

Finally, the time series were reduced to a within group average and an independent ttest p-value in terms of time (see figures). The p-value indicates the probability of false positives when the null hypothesis stating that the two groups are identical is true. This was mainly computed to integrate variance information.

To allow a better visualization and simplify interpretation of the results, the previously mentioned time series were low pass filtered with a moving average of 5 days for actors products' sales metric and for their p-value estimate.

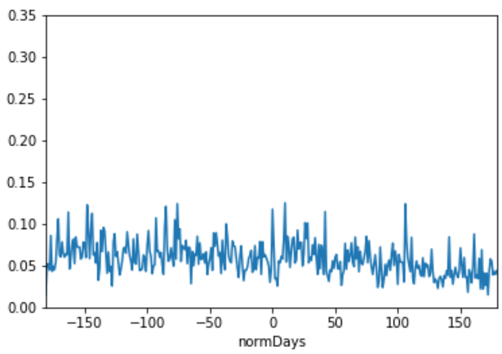


Figure 2: Evolution of popularity/amount of sales metric in time for actors controls

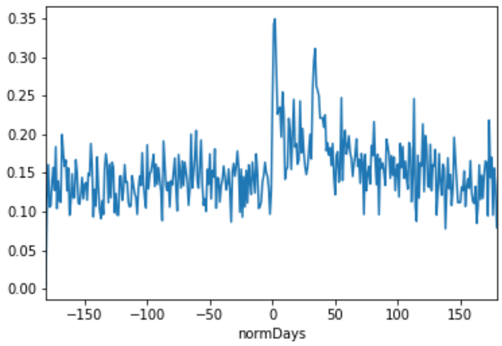


Figure 3: Evolution of popularity/amount of sales metric in time for actors subjects

# 4 Results

## 4.1 Actors

The actors retained as popular, is that to say with a threshold of at least 100 reviews count in the required time window, were only 90 over 500, which represents 18% of the original subjects.

As can be seen on figures 3, there is an important increase of the metric from 0.12 to 0.35 (191 % increase) at the time of death, which is zero on the x-axis. This inscrease is not visible for controls (figure 2) whose values stay between 0.05 and 0.10.

Note however, that the rest value for controls is slightly shifted compared to subjects (in average -0.04). This may come from an imperfect matching, probably more pronounced for famous subjects. However the matching was done mostly to have comparable variations in review time series before the death event. From this point of view the two series are well matched.

The p-value decreases also drastically after the

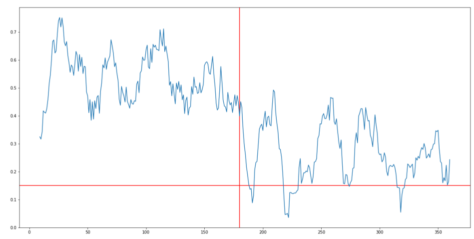

Figure 4: Independent ttest p-value in time for actors

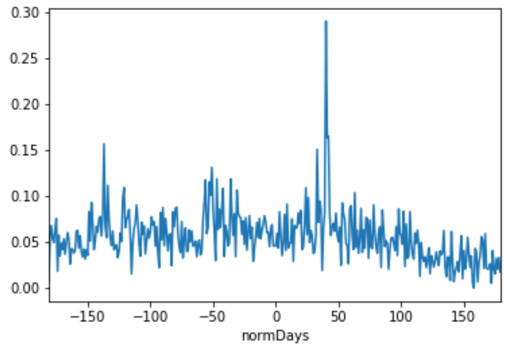


Figure 5: Evolution of popularity/amount of sales metric in time for musicians controls

death day (represented by the vertical red line) and crosses a threshold of 0.15 a few days after the event. This is still a high p-value, indicating that the within-group variances may be high and care has to be taken regarding the results.

## 4.2 Musicians

Here also an important incease is visible at the time of death for the subjects (366 %) in figure 6. However the within group variance seems high and there is a peak appearing close to 50 days post-zero event for the controls (roughly 150 % increase, figure 5). Nevertheless the p-value decreases drastically after the death event and reaches a significative threshold of less than 0.05 (see figure 7).

# 5 Conclusion

As a general conclusion, It seems that for some proportion of actors and musicians, there is visible effect related to death. Indeed, the curves of reviews count for the subjects show an increase of roughly 200% for actors and 366% after the death day, which is an important increase. However, the p-value for actors, used to test for dif-

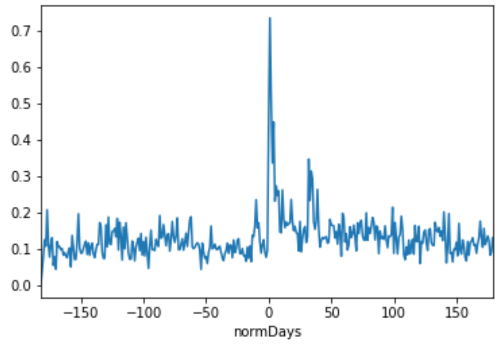


Figure 6: Evolution of popularity/amount of sales metric in time for musicians subjects

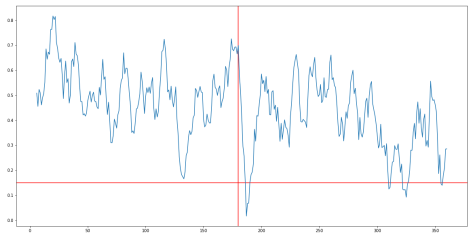

Figure 7: Independent ttest p-value in time for muscic interpets

ferences between subjects and controls does not reach a significative threshold, even If it decreases drastically after the time of death. This shows that the within group variance is high and probably, the proportion of actors' subjects that have a significative increase is more restricted. In the case of musicians the p-value reaches a significative level just after the death day, meaning that probably the effect of post-mortem artwork sales increase is more marked for musicians. This is not surprising since music products are more intuitively associated with the corresponding artist than films with the actors playing on it. Moreover, the data of music interprets contained really famous icons such as Mickeal Jackson, which don't have an equivalent in terms of fame in the actors' group.

## References

@misctitle=cnbc.comfive-artists-who-saw-a-boom-in-music-sales-after-death.html,url=`https://www.cnbc.com/2017/07/28/five-artists-who-saw-a-boom-in-music-sales-after-death.html`,

@misctitle=billboard.comlinkin-park-sales-rise-chester-bennington-death,url=`https://www.billboard.com/articles/columns/chart-beat/7874023/linkin-park-sales-rise-chester-bennington-de`