Data Science and IT Project Management

# Project Management

### Introduction

#### **Main goal of the project**

While searching for interesting data sets, we found the Young People Survey. The survey was conducted 2013 by students of the statistics class from the Culty of Social and Economic Sciences of the Comenius University in Bratislava. The participants got the poll in written form, as well as in electronic form. As the University is in Slovakia, the language of the survey was Slovak, but it was translated afterwards into English. Likewise, all the participants were Slovaks. Because mostly students filled out the questionnaire, all participants were aged between 15 and 30.

While studying the dataset we all pretty fast asked ourselves, if it was possible to predict all of the user’s interests just by knowing some of their interests. That’s why we constructed our main goal, to find out whether it is possible to predict a person’s music taste based on their movie taste. The same case reversed is also examined, just like the question, if it is possible to predict the individual taste of a person in general. The data is rated data, but we treat them like qualitative data as we assume that the distance between the rating 1 and 2 is the same between for example 4 and 5.

#### Costs of the IT Project

As we are students and doing this project for our course, monetary costs will not be an issue. But talking about costs and resources in general, the time and the workload might cause problems. We only have a limited time frame to work, so the workload should not be too extended.

#### Time constraints

Each project member has 15 hours to work on the project during class lessons, in addition, we are planning to spend about 10 hours per person outside of the lecture.

#### Usefulness of the project

The projects aim is to answer the question, if it is possible to predict music taste based on move tastes. In addition, we want to evaluate if this is possible vice versa, and in general, if it is possible to predict one’s individual taste.

#### Maintenance

To use the results of the IT project on long term, the datasets have to be updated regularly, as they get outdated otherwise. In addition, the whole project has to be relaunched after that certain amount of time, to keep the conclusions current. Due to many circumstances it is possible that the results from the data analytics change, which means that the accuracy of our results is only assured in our specific framework.

#### Our customers

The costumers of our report are the readers of our results. Everyone who is interested if there is a correlation between any kind of music and movies is our costumer.

#### Group and Project Management

To develop a great result, we needed to find the best strategy for our Project Management. Our main objectives were to be fast but still simple, as we had a short time to deliver value and high collaboration and communication, because we had different experts for the topics in our team which should work on their own, but also keep the team updated and prevent misunderstandings.

Therefore, we decided to use the Scrum-model. In our particular situation it is not possible to use SCRUM in the commonplace way, as we are a very small team doing a small project. Anyway, we want to use the concepts, we just tailored them to our situation.

We planned two sprints, first one 14th of May until 11th of June and second Sprint from 12th of June until 15th of July. Therefore, we also have two Sprint Planning and Sprint Review Meetings.

Instead of doing a daily scrum, we do a meeting likewise right before every Data Science/IT Project Management course. Those meetings improve communications, eliminate other meetings, identify impediments to development for removal, highlight and promote quick decision-making, and improve the Development Team’s level of knowledge.

Our task is to assure Maintenance, Planning and Organizing. We do that with our Sprint Planning Meetings, where we set our plan for the following sprint. We also must assure Estimating, Scheduling and Monitoring – we do that with our Daily Sprint Meetings. To be aware of risk and quality, we have a Product Owner.

Because there is no real Product Owner, we decided to appoint Clara to be our PO, in a slightly different way. She is the person that reassures that the team achieves the goals and missions, she knows the requirements and keeps close contact to our customers, therefore she is the person that keeps up with the Product Backlog, including its content, availability, and ordering. Furthermore, she will verify the correlation results from the development team.

Hanna is our Scrum Master. Her task is to coach the team in self-organization and to improve communication. In addition, she must remove impediments to the development team’s progress and organize the Scrum events. She is the one responsible for the assignment writing, presentation and organization with for example the excel file during the process.

The development team consists of Jana and Anna. They work self-organized, while creating complex code to turn the Sprint Backlog into working functionality. The sprint backlog is a list of tasks identified by the Scrum team to be completed during the Scrum sprint. During the sprint planning meeting, the team selects some number of product backlog items and identifies the tasks necessary to complete them. They will code with R to identify possible correlations regarding our data.

#### Milestones

Milestones are a typical characteristic from a linear sequential IT project model, like the waterfall concept. It’s not possible to map traditional, waterfall project management concepts into a Scrum/Agile framework, because Scrum and Agile project management are fundamentally different from traditional project management in that the process is specifically designed to be loosely-coupled to the objectives being pursued. As we still need milestones, we decided to use each point from the Product Backlog as a milestone, which can be found at “II.Sprint – Product Backlog”.

### 1. Sprint

#### Product Backlog

|  |  |  |  |
| --- | --- | --- | --- |
| Task Name | Priority | Status | Assigned to Sprint |
| team organisation | High | Complete | Yes |
| data research and collection | High | In progress | Yes |
| formulating question and hypothesis | Medium | In progress | Yes |
| project management with excel/protocoll | High | In progress | Yes |
| data cleaning | Medium | In progress | Yes |
| correlation | High | Not started | No |
| correlation analysis | High | Not started | No |
| interpret findings | Medium | Not started | No |
| writing protocol/analysis | Medium | Not started | No |
| findings into figures/text | Low | Not started | No |
| merge findings | Low | Not started | No |
| design the presentation | Medium | Not started | No |
| complete presentation+paper | High | Not started | No |

We use an Excel for our Project Management. We made a timetable to show how we are planning our sprints, and a table for our Product Backlog.

1 Product Backlog in the first Sprint Planning Meeting

#### Sprint Planning Meeting 23.05.2018

The Scrum Master facilitates the meeting, while the product owner presented his vision to the team and explained the requirements.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Task Name | Responsible | Start | End | Days | Status |
| Sprint 1 |  | 2018-05-23 | 2018-06-05 | 13 |  |
| team organisation | Hanna | 2018-05-23 | 2018-05-24 | 1 | Complete |
| data research and collection | Clara | 2018-05-24 | 2018-05-27 | 3 | In progress |
| formulating question and hypothesis | Clara | 2018-05-27 | 2018-06-03 | 7 | In progress |
| project management with excel/protocoll | Hanna | 2018-05-28 | 2018-06-04 | 7 | In progress |
| data cleaning | Clara | 2018-06-03 | 2018-06-05 | 2 | In progress |

2 1.Sprint Backlog

3 Sprint Backlog for 1.Sprint

In addition, he provides information about the product backlog. Afterwards, the team estimates what can be done in the first sprint, they define the goals with the product owner. Now, the team knows what to do. Because of this Meeting the first point, to organize the team, is already done.

#### Daily Scrum 25.05.2018

The daily scrum meeting is not used as a problem-solving or issue resolution meeting, the meeting takes place to identify issues and impediments. During the daily scrum, each team member answers the following three questions:

What did you do yesterday?

What will you do today?

Are there any impediments in your way?

Because this was our first Daily, it took a bit longer than 15 minutes, but we will improve that. Clara and Hanna informed the team about their data research regarding the Young people survey. There were no impediments.

#### Daily Scrum 30.05.2018

Data research and collection task are finished, we started with the Project Management protocol and designed a dedicated excel file for our project. We discussed and consolidated a hypothesis and question regarding our analysis. The question we are investigating is whether young peoples taste in music correlates with their taste in movies.

#### Daily Scrum 01.06.2018

Today was the first time we discovered an issue. We wanted to start continuing our protocol but weren’t sure how much into detail we have to describe our Project Management process. Hanna will take care of the question, as she is the Scrum Master and Clara as the product owner will find out about the requirements regarding that topic.

#### Sprint Review Meeting and Sprint Retrospective 04.06.2018

Sprint Review is held at the end of the Sprint to review the progress and change the Product Backlog if needed. Sprint Retrospective is a meeting for the team, so that they can discuss what went well, what to improve and to make actionable commitments. During that meeting the Scrum Team collaborate about what was done in the Sprint. They adapt changes to the Product Backlog and communicate about what to improve. Basically, feedback should be collected, and collaboration should be fostered.

The Scrum Master ensures that the event takes place and that attendees understand its purpose; the duration of the meeting takes about an hour.

The Product Owner, Clara, explains what Product Backlog items have been “Done” and what has not been “Done” – thankfully in our case, every item is “Done”. The Development Team discusses what went well during the Sprint, what problems it ran into, and how those problems were solved. Because there was not much developing work, everything went quite well.

The entire group collaborates on what to do improve, and we decided to improve our preparation for the Daily Sprint Meetings, so that we really tell our team members everything that is important for them.

Because this meeting should also be a review of the timeline, budget, potential capabilities, and marketplace, we discussed if anything is jeopardized. We found out that the time may be an issue, so we decided to improve our time management and work more efficient.

The result of the Sprint Review is the same Product Backlog because we didn’t have to make any changes.

### 2. Sprint

#### Sprint Planning Meeting 05.06.2018

Because this Sprint is our last one, we must put all of the remaining Backlog items on our Sprint Backlog.

The meeting starts with reviewing our first sprint, which was more about organizing, defining and controlling the project. Because today is also a day that a Daily Scrum should take place, we decided to put them both together. We achieved many milestones, we even managed to be in time with all our Scrum Backlog items. Now the real IT-work is going to happen. Therefore, it is very important to remind the team of the big picture, and what we are aiming for. In addition, the Scrum Master calls for a group consensus on the plan and Team with Product Owner signal, if this is the best plan they can make given what they know right now. Our plan is to start with the correlation, continue with interpretation and understanding our results, and parallel always work on the paper and the presentation. After every team member confessed to the planned sprint backlog and time table, everybody knew what to work on next.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sprint 2 | Responsible | 2018-06-05 | 2018-06-22 | 17 |  |
| correlation | Jana | 2018-06-05 | 2018-06-07 | 2 | Not started |
| correlation analysis | Anna | 2018-06-07 | 2018-06-10 | 3 | Not started |
| correlation verification SPSS | Clara | 2018-06-08 | 2018-06-11 | 3 | Not started |
| interpret findings | Anna | 2018-06-11 | 2018-06-16 | 5 | Not started |
| writing assignment | Hanna | 2018-06-10 | 2018-06-12 | 2 | Not started |
| writing analysis | Jana | 2018-06-11 | 2018-06-13 | 2 | Not started |
| findings into figures/text | Jana | 2018-06-16 | 2018-06-20 | 4 | Not started |
| merge findings | Anna | 2018-06-17 | 2018-06-19 | 2 | Not started |
| design the presentation | Hanna | 2018-06-19 | 2018-06-21 | 2 | Not started |
| complete presentation+paper | All | 2018-06-21 | 2018-06-22 | 1 | Not started |

4 2.Sprint Backlog

5 time chart for sprints

#### Daily Scrum 06.06.2018

The correlation with the programming statistical language R causes some problems, as it is not clear how to execute some necessary commands. It seems like we will not be in our time schedule regarding that issue, because we will need more time. Anyway, we plan to remove stumbling block. We will double check our results with SPSS.

#### Daily Scrum 13.06.2018

Because the correlation itself caused some problems, we couldn’t analyze the correlation in time. Also, to verify our correlation, we did an additional T-test. We plan to do some overtime and first analyze the correlation on ourselves, afterwards we will discuss our results so that we may find more interesting outcomes.

#### Daily Scrum 15.06.2018

As we worked a lot in the last two days, a lot of work has been done. The protocol is almost finished, and the correlation analysis combined from all our results satisfies all of the team members. We agreed on what team member will design which part of the presentation and who’s the one who will merge them together. Also, the text protocol must be finished, and some figures should be included. We agreed on putting our Scrum Excel, which we continually updated during our Daily Meetings, in our assignment.

#### Daily Scrum 20.06.2018

Not only the presentation, but also the assignment with its textual and figuratively part is finished. We decided that everybody must give the presentation and the assignment a last checking look, and afterwards Clara will send it to our professor. In addition, we scheduled a meeting for the afternoon, so that we can practice our presentation together.

#### Sprint Review and Sprint Retrospective Meeting 22.06.2018