



AVALIAÇÃO DE BASES DE DADOS DE MASSIVE OPEN ONLINE COURSES (MOOCS)

RELATÓRIO FINAL DE PROJETO DE INICIAÇÃO CIENTÍFICA (PIBIC/INPE/CNPq)

Vanessa Cordeiro Gonçalves de Oliveira (INPE, Bolsista PIBIC/CNPq)
E-mail: vanessa.cordeiro148@gmail.com

Eduardo Martins Guerra (CTE/LAC/INPE, Orientador)
E-mail: eduardo.guerra@inpe.br

COLABORADORES

Dr. Clovis Torres Fernandes (ITA)
Dr. Fabio Kon (IME/USP)
Dr. Paulo Lemos (Unicamp)

Julho de 2018



SUMÁRIO

LISTA DE SIGLAS	3
LISTA DE FIGURAS	4
RESUMO	5
1. INTRODUÇÃO	6
2. OBJETIVO	6
2.1. <i>Objetivo Geral</i>	6
2.2. <i>Objetivos Específicos</i>	7
3. FUNDAMENTAÇÃO TEÓRICA	7
4. MÉTODOLOGIAS	8
5. RESULTADOS	11
6. CONCLUSÃO	15
REFERÊNCIAS BIBLIOGRÁFICAS	16



LISTA DE SIGLAS

CSV: *Comma-separated values*.

INPE: Instituto Nacional de Pesquisas Espaciais.

ITA: Instituto Tecnológico de Aeronáutica.

MOOC: *Massive Open Online Course*, do português “Curso Online Aberto e Massivo”.

TDD: *Test-Driven Development*, do português “Desenvolvimento orientado a teste”.

UNICAMP: Universidade Estadual de Campinas.

USP: Universidade de São Paulo.



LISTAS DE FIGURAS

Figura 1: Students who completed and those who started, but did not complete the activities.	11
Figura 2: Percentage of students who started and did not complete.	11
Figura 3: Percentage of students who started and did not complete (only lessons with video).	12
Figura 4: Number of students who did the activity as their last activity in the course.	13
Figura 5: Average number of times students did an activity.	13
Figura 6: Percentage of students who did the activity more than once.	14
Figura 7: Weeks students took to complete the course.	14



RESUMO

Os *Massive Online Open Courses* (sigla MOOCs) vem ganhando cada vez mais popularidade. São cursos online e abertos, criados com o objetivo de receber participações em larga escala. Para a criação de cursos nesse formato em língua portuguesa, a plataforma Coursera firmou parceria com instituições brasileiras de ensino e passou a oferecê-los. Como resultado, alunos e profissionais de diversos países de língua portuguesa tiveram a oportunidade de participar de cursos elaborados por essas universidades. Este trabalho, iniciado em agosto de 2017, tem como objetivo realizar uma análise nos dados gerados pelos MOOCs dessas instituições, podendo assim investigar o comportamento dos alunos dentro do curso, o ponto em que os que não completam costumam desistir com maior frequência e como é o desempenho deles para cada tarefa, a fim de se levantar os perfis de comportamento dos estudantes que participam desses tipos cursos. Os cursos que tiveram seus dados analisados são: TDD – Desenvolvimento de Software Guiado por Testes, Princípios de Desenvolvimento Ágil de Software, Orientação a Objetos com Java, Empreendedorismo e as Competências do Empreendedor, Desenvolvimento Ágil com Java Avançado, Introdução à Ciência da Computação com Python Parte 1 e Parte 2. Inicialmente, foi realizado um estudo sobre dados que seriam disponibilizados pela Coursera para a elaboração de objetivos. Após esta etapa, partiu-se para a importação e a criação de uma base de dados para cada um dos cursos. Com as bases de dados já estabelecidas, iniciou-se a realização de consultas e análise de dados com a intenção de extrair as seguintes informações: quantidade de alunos que concluíram cada atividade e os que começaram mas não concluíram as mesmas; quantidade de alunos que abandonaram o curso em cada tarefa; quantidade de vezes que um aluno fez uma atividade; percentual de alunos que iniciaram e não concluíram cada atividade; percentual de alunos que fizeram uma determinada atividade mais de uma vez; e média de vezes que um aluno fez uma atividade. Tendo os dados dessas novas informações disponíveis em tabelas, foram elaborados gráficos para que essas informações pudessem ficar mais visíveis e compreensíveis. Já com os gráficos, avaliações foram feitas, as primeiras conclusões surgiram, e assim, novos objetivos foram traçados, sendo eles a revisão dos gráficos até então disponíveis e a criação de histogramas para observar o tempo comum em que os alunos geralmente demoram para concluir os cursos. Agora, depois da conclusão das avaliações e já tendo informações sólidas e concretas, o artigo a respeito do trabalho sofreu um avanço maior e já está quase sendo concluído.

1. INTRODUÇÃO

Nos dias de hoje, é impossível negar a importância da internet na sociedade. Cada vez mais, podemos observar como essa ferramenta - cada vez mais democrática - está presente na vida cotidiana das pessoas. Seja na forma em que elas se relacionam entre si ou até no modo como elas buscam adquirir informações.

É em decorrência dessa importância, somada a necessidade de oferecer educação de qualidade ao máximo possível de pessoas, que se passou a pensar na internet como um meio de disseminar conhecimento de forma simples e conveniente para os usuários. Afinal, como se pode quebrar barreiras geográficas e sociais, e oferecer os mais diversos tipos de ensino a pessoas que antes não tinham acesso aos mesmos?

Foi então, a partir dessa ideia, que surgiram os Massive Online Open Courses (da sigla MOOCs). Como o próprio termo sugere, trata-se de cursos online e abertos criados com o objetivo de receber participações em larga escala.

Esses cursos começaram a ter grande popularidade a partir de 2012, e desde então vem crescendo cada vez mais o número de adeptos e plataformas dedicadas aos mesmos, sendo a Coursera uma dessas plataformas.

Fundada em 2012, a Coursera oferece cursos de diversas universidades renomadas no mundo. Os cursos são disponibilizados no formato on-demand ou em períodos específicos (a critério da instituição que ministra o curso), com aulas em vídeos e uma vasta gama de opções de leitura obrigatória e extra. Além disso, a plataforma também conta com um aplicativo para IOS e Android, que possibilita gravar os conteúdos para acesso offline.

Em 2015, a Coursera firmou uma parceria com instituições brasileiras (COURSEERA BLOG, 2015), como o ITA, USP e UNICAMP, para a criação de cursos na língua portuguesa. Como resultado, alunos e profissionais de diversas partes do país, e até mesmo de outros países de língua portuguesa, tiveram a oportunidade de participar de cursos ministrados por essas universidades.

2. OBJETIVOS

2.1. OBJETIVOS GERAIS

O objetivo geral deste trabalho é fazer uma análise nos dados gerados pelos cursos das instituições brasileiras na Coursera, a fim de estudar o perfil e o comportamento dos estudantes que participam desse tipo de curso.

Os dados que serão utilizados para realizar essa análise são disponibilizados pela própria plataforma, e através deles será possível traçar um comportamento dos alunos



dentro do curso, o ponto em que os que não completam costumam desistir com maior frequência e como é o desempenho deles para cada tarefa.

2.2. OBJETIVOS ESPECÍFICOS

Após realizar as análises e tirar conclusões sobre o perfil e comportamento dos alunos em cursos do tipo MOOC da Coursera, um artigo será escrito e publicado pelos professores Eduardo Guerra e Fabio Kon com o intuito de contribuir no desenvolvimento dos cursos MOOC.

3. FUNDAMENTAÇÃO TEÓRICA

Nos últimos, vem crescendo cada vez mais o número de adeptos e plataformas dedicadas aos chamados MOOCs. Organizações como a Coursera, edX, Udacity e outras, vem cada mais ganhando adeptos e aumentando seu catálogo de cursos através de parcerias com instituições de ensino ou empresas de tecnologia que investem em educação.

Em um artigo do “*New York Times*”, com o título “*The Year of the MOOC*”, LAURA PAPPANO (2012) escreveu sobre o grande “*boom*” que plataformas que oferecem cursos do tipo MOOCs tiveram naquele ano e a importância e desafios que as mesmas vinham assumindo dali em diante.

A mesma autora cita os números surpreendentes que algumas plataformas tiveram em seus primeiros meses e como isso pegou de surpresa os responsáveis (funcionários, presidentes, fundadores e etc) por essas plataformas. Ela cita que

"A pintura está quase seca, mas a edX, empresa iniciante sem fins lucrativos de Harvard e do Instituto de Tecnologia de Massachusetts, tem 370 mil estudantes neste outono em seus primeiros cursos oficiais. Isso não é nada. O Coursera, fundado em janeiro passado, alcançou mais de 1,7 milhões - crescendo “mais rápido que o Facebook”, conta Andrew Ng [...] “Isso nos pegou de surpresa”, diz David Stavens, que formou uma empresa chamada Udacity com Sebastian Thrun e Michael Sokolsky depois que mais de 150.000 pessoas se inscreveram para a “Introdução à Inteligência Artificial” do Dr. Thrun no último outono [...]” (LAURA PAPPANO; NYT, 2012, tradução nossa) ¹

A autora supracitada também levanta pontos que devem ser discutidos em relação aos MOOCs. Um deles é que por serem direcionados a muitos alunos, nem sempre permitem que os professores atendam de forma individual cada aluno. Outro ponto importante citado, é a questão de avaliação, já que podemos considerar que possa ocorrer cópia de trabalhos e colas nestes cursos.

¹ “The paint is barely dry, yet edX, the nonprofit start-up from Harvard and the Massachusetts Institute of Technology, has 370,000 students this fall in its first official courses. That’s nothing. Coursera, founded just last January, has reached more than 1.7 million — growing “faster than Facebook,” boasts Andrew Ng [...] “This has caught all of us by surprise,” says David Stavens, who formed a company called Udacity with Sebastian Thrun and Michael Sokolsky after more than 150,000 signed up for Dr. Thrun’s “Introduction to Artificial Intelligence” last fall [...]”

Uma crítica que autora supracitada faz é o grande número de evasões que esses cursos geralmente tem. Essas evasões podem ser associadas a diversos tipos de problemas, sendo alguns deles: falta de disciplina do aluno, falta de autonomia, ou até falta do pré-requisito do aluno para realizar o curso.

Mas fato é, que apesar ainda se ter muito a que pesquisar e melhorar, as já consolidadas plataformas que oferecem esses tipos de curso, cresceram e veem crescendo cada vez mais. Inclusive, abrindo espaço para novas plataformas (com ideias e outras variedades de cursos) surgirem.

A Coursera (plataforma que teve seus dados estudados neste projeto de IC) foi criada em 2012, e em 2015 (um pouco mais de três anos após sua fundação) já contava com mais de 12 milhões de usuários matriculados em mais de 1.000 cursos, que eram disponibilizados na época (KNOWLEDGE@WHARTON, 2015).

Hoje, a Coursera já ultrapassou a barreira dos 31 milhões de usuários em todo planeta (ESTADÃO, 2018), e oferece atualmente mais de 2.902 cursos, em parceria com 171 instituições (entre universidades e empresas) de 28 países (COURSERA, 2018).

O crescimento da Coursera e de outras plataformas que oferecem MOOCs é constante, e é de se esperar que devido a facilidade como os mesmos são oferecidos, a variedade dos mesmos, e por serem muitas vezes ministrados por grandes instituições, cada mais pessoas passem a aderir a esse meio de ensino. Como cita CLÁUDIA e ANA PAULA (2013)

“[...] O que não se pode negar é que a barreira entre on-line e presencial nas universidades tende a ficar cada vez mais indefinida, em especial porque futuros alunos, alunos e ex-alunos buscam nas universidades conhecimentos em ampla variedade de assuntos e a tendência mundial é que tenham seu acervo de conhecimento à disposição.” (CLÁUDIA; ANA PAULA, 2012)

4. MÉTODOLOGIAS

No início do projeto, alguns procedimentos metodológicos foram propostos. A seguir, os procedimentos são descritos:

- **Seleção de cursos para análise e obtenção dos dados do Coursera.**

Primeiramente, foi realizado a seleção dos cursos da Coursera que teriam seus dados analisados. Foi acertado que o projeto seria desenvolvido com base nos dados de alguns cursos das seguintes universidades brasileiras: Universidade de São Paulo (USP), Universidade Estadual de Campinas (UNICAMP) e Instituto Tecnológico de Aeronáutica (ITA).

Os cursos que teriam seus dados analisados eram:

- TDD – Desenvolvimento de Software Guiado por Testes - ITA;
- Princípios de Desenvolvimento Ágil de Software - ITA;



- Orientação a Objetos com Java - ITA;
- Empreendedorismo e as Competências do Empreendedor - UNICAMP;
- Desenvolvimento Ágil com Java Avançado - ITA;
- Introdução à Ciência da Computação com Python Parte 1 - USP; e
- Introdução à Ciência da Computação com Python Parte 2 - USP.

A Coursera, através da plataforma, oferece aos responsáveis pelo curso o acesso a arquivos CSV², que contém diversos dados referente ao desempenho dos estudantes dentro do curso, como, por exemplo, o horário que um determinado aluno concluiu uma atividade ou quais alunos concluíram o curso.

Vale destacar, que nesses arquivos não contém nenhum tipo de dados pessoais dos estudantes (como nome, RG, CPF, e-mail e etc). Os estudantes são identificados por chaves ID³ que são geradas aleatoriamente pela Coursera.

Para ter acesso aos dados desses cursos, foi realizado o contato com os professores responsáveis por ministrar os mesmos. Cada um deles enviou esses arquivos para contribuir com o projeto.

- **Criação das bases de dados, importação dos dados e análise do modelo de dados do Coursera.**

Após ter acesso aos arquivos CSVs disponibilizados pelos professores, foi realizado a criação das bases de dados de cada curso e a importação dos dados dos arquivos CSVs.

Utilizando o sistema de gerenciamento de banco de dados, MySQL, para cada curso foi criado um banco de dados. Em cada um desses bancos foi realizado a importação dos dados dos CSVs para as suas tabelas.

Com as bases de dados prontas, foi realizado uma análise do modelo de dados da Coursera com o objetivo de conhecer as bases de dados, poder observar os tipos de dados que elas trazem e quais os tipos de informações dali podem ser extraídas.

- **Criação das consultas para obtenção das informações sumarizadas dos MOOCs, elaboração de gráficos e conclusões a respeito.**

No início de outubro de 2017, com as bases de dados já estabelecidas e estudadas, iniciou-se a criação de consultas para extrair informações que pudessem colaborar com a análise de perfil e do comportamento dos alunos.

² Comma-separated values, da sigla CSV, é um formato de arquivo que armazena dados tabelados. São arquivos de texto separados por um delimitador, que usam a vírgula e a quebra de linha para separar os valores.

³ ID, o nome de usuário ou e-mail ou código que servem de identificação pessoal aos internautas, enquanto acessam certos sites (sobretudo clientes de e-mail e/ou acesso à internet).



Buscou-se extrair as seguintes informações das bases de dados:

- a. Quantidade de alunos que concluíram cada atividade e de alunos começaram, mas não concluíram cada atividade.
- b. Percentual de alunos que iniciaram e não concluíram em cada atividade.
- c. Por curso: quantos alunos vieram de uma atividade imediatamente anterior, quantos vieram de uma bem anterior e quantos vieram de uma posterior.
- d. Quantidade de alunos que abandonaram o curso em cada tarefa.
- e. Por curso: média de vezes que os alunos repetiram uma atividade.
- f. Porcentagem de alunos que fez uma determinada atividade mais de uma vez.

Enquanto essas informações eram extraídas, gráficos eram gerados e os mesmos eram avaliados, e quando se encontravam inconstâncias, as consultas eram refeitas e os gráficos reconstruídos.

Esse processo de extração de informações, e construção e reconstrução de gráficos foi constantemente repetido para que fossem adquiridos dados e informações mais precisos possíveis e de máxima qualidade.

Também foi solicitado a criação de histogramas para indicar o tempo comum (em semanas) que os alunos demoram para concluir cada um dos cursos. Essas informações foram extraídas, avaliadas e os histogramas foram elaborados.

Ao decorrer que os gráficos e histogramas iam ficando prontos, os professores Eduardo Guerra e Fabio Kon iam discutindo e tirando conclusões sobre os mesmos, afim de utilizar essas informações no artigo (que será escrito em inglês) sobre o perfil e comportamento dos alunos.

- **Escrita de artigo a respeito do trabalho desenvolvido.**

Tendo os gráficos e histogramas “em mãos”, os professores Eduardo Guerra e Fabio Kon tiveram a oportunidade avançar mais ainda na escrita do artigo.

Por volta do início de julho de 2018, já havia um grande avanço na elaboração do artigo e faltava pouco para a conclusão do mesmo. Sendo que ainda era necessário a conclusão de alguns pontos importantes levantados por eles, além da revisão de todo o artigo.

5. RESULTADOS

Após meses que se passaram realizando consultas para a extração de informações, foram gerados alguns gráficos e histogramas para visualizar essas informações.

A seguir, os gráficos e o histograma referente ao curso de “Introdução a Ciência da Computação com Python - Parte 1” poderão ser observados, como exemplo dos resultados que se obteve:

a. Quantidade de alunos que concluíram cada atividade e de alunos começaram, mas não concluíram cada atividade.

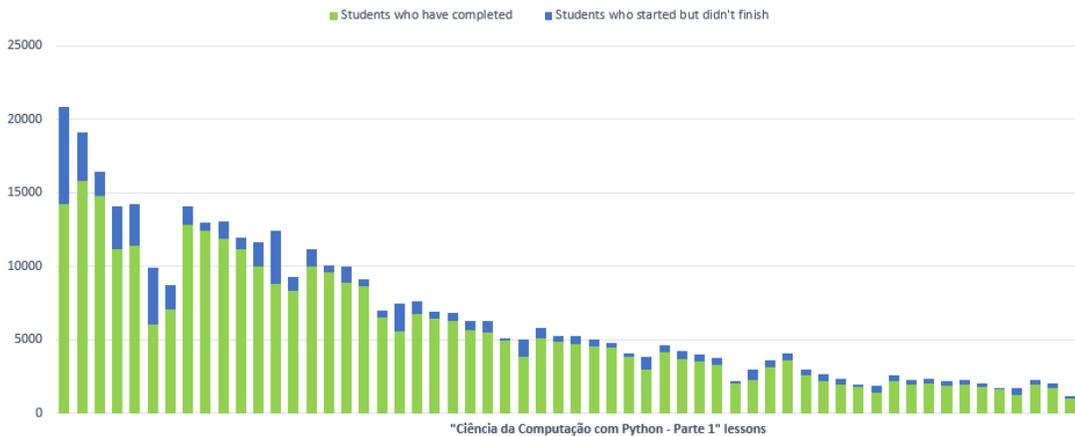


Figura 1 - Students who completed and those who started, but did not complete the activities.

Esse gráfico mostra a quantidade de alunos que concluíram cada atividade (em verde) e a quantidade de alunos que começaram, mas não concluíram cada uma das atividades (em azul). O acumulo das duas barras geram o total de alunos que cursaram o curso (independente se concluíram ou não).

O intuito é mostrar a relação da quantidade que concluem versus a quantidade que iniciam o curso, mas não terminam, em relação ao total de alunos que iniciam o curso.

b. Percentual de alunos que iniciaram e não concluíram em cada atividade.

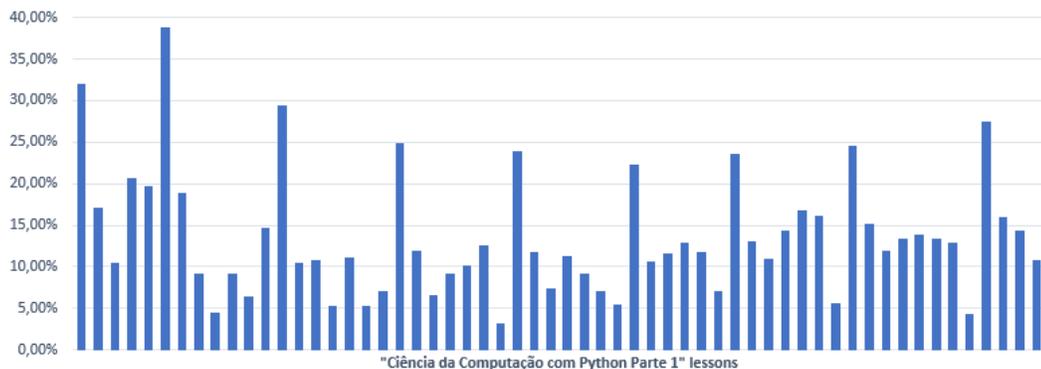


Figura 2 - Percentage of students who started and did not complete.

Esse gráfico mostra o percentual de alunos que iniciaram, mas não concluíram cada atividade. A ideia do gráfico é dar uma noção de quais atividades são mais abandonadas antes de serem concluídas.

Além desta versão, foi feita outra versão contento apenas as atividades com vídeos. Nessa versão, as atividades foram ordenadas pelo tempo dos vídeos, do maior para o menor.

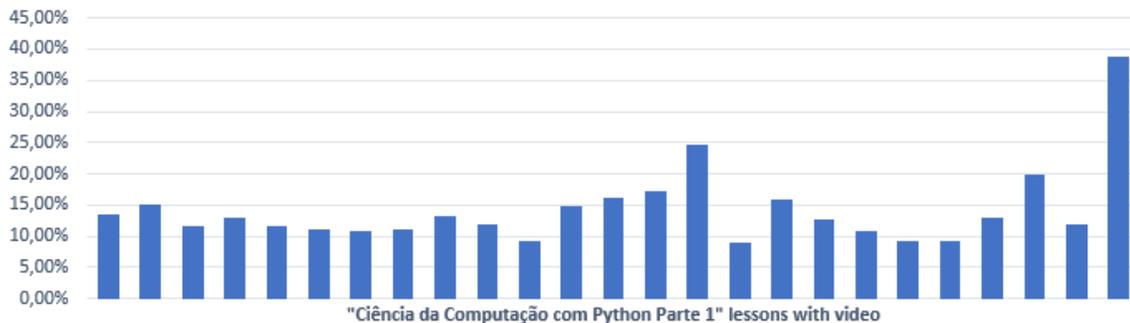


Figura 3 - Percentage of students who started and did not complete (only lessons with video).

O objetivo dessa versão é tentar mostrar se o tempo do vídeo pode influenciar o aluno a abandonar a atividade.

c. Por curso: quantos alunos vieram de uma atividade imediatamente anterior, quantos vieram de uma bem anterior e quantos vieram de uma posterior.

Ocorreu um certo atraso na elaboração desses gráficos e o de “Introdução a Ciência da Computação – Parte 1” acabou não ficando pronto a tempo até a entrega do Relatório Final.

Esse gráfico está sendo feito por curso e consiste em 3 barras, onde na primeira barra será exposto quantos alunos vieram de uma atividade imediatamente anterior; na segunda, a quantidade de alunos que vieram de uma bem anterior; e na terceira, a quantidade de alunos que vieram de uma posterior.

A ideia é mostrar como os alunos se comportam no curso em relação a ordem das atividades. Se eles seguem as atividades na ordem proposta pelos professores, se pulam algumas atividades ou se tem o comportamento frequente de voltar e rever uma atividade anterior a que eles se encontram.

d. Quantidade de alunos que abandonaram o curso em cada tarefa.

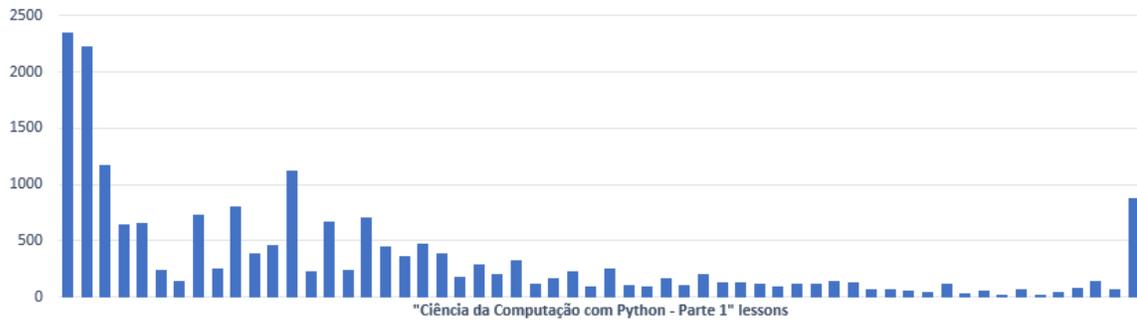


Figura 4 - Number of students who did the activity as their last activity in the course.

Esse gráfico mostra a quantidade de alunos que abandonaram o curso em cada tarefa.

Sendo o intuito mostrar em quais atividades os alunos fazem por último antes de abandonar o curso em definitivo.

e. Por curso: Média de vezes que os alunos repetiram uma atividade.



Figura 5 - Average number of times students did an activity.

Esse gráfico mostra a média de vezes que os alunos repetiram cada atividade.

O intuito é demonstrar a média de vezes que os alunos costumam repetir uma determinada atividade.

f. Porcentagem de alunos que fez uma determinada atividade mais de uma vez.

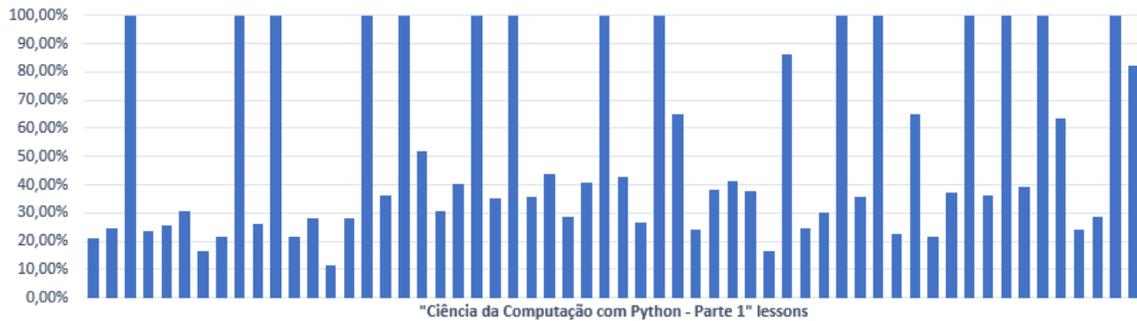


Figura 6 - Percentage of students who did the activity more than once.

Esse gráfico mostra o percentual de alunos que fizeram uma atividade mais de uma vez.

O intuito é mostrar em relação ao total de alunos que iniciaram cada atividade, o percentual de alunos que chegaram a repetir as mesmas.

g. Histograma: o tempo comum (em semanas) que os alunos demoram para concluir cada um dos cursos.

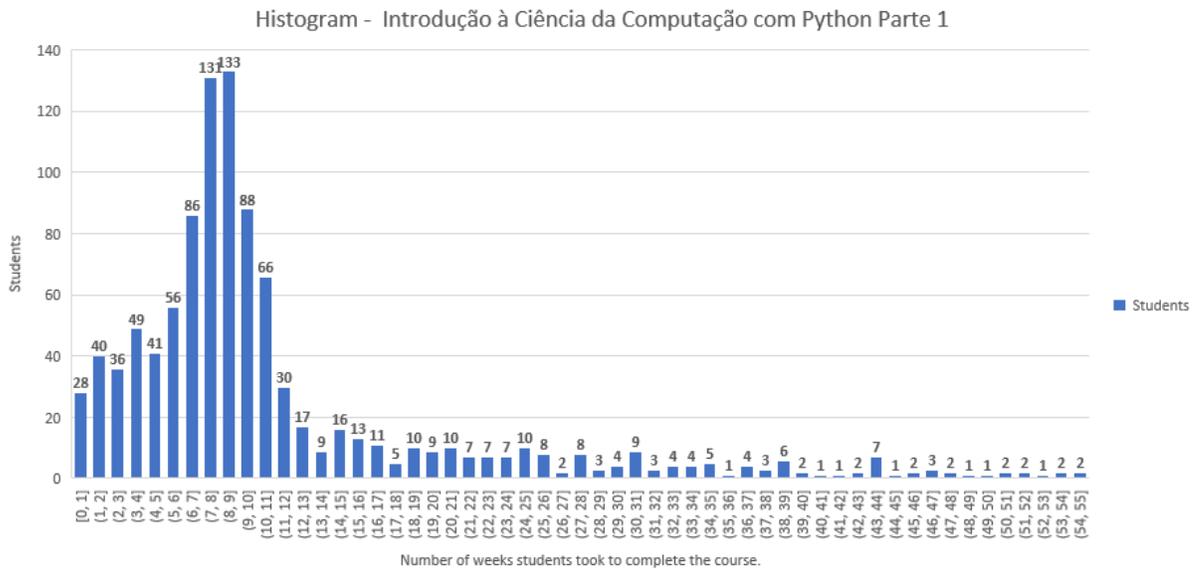


Figura 7 - Weeks students took to complete the course.

Esse histograma mostra o tempo comum que os alunos demoram para concluir cada um dos cursos.

Com amostras de uma semana, o intuito é mostrar a frequência que os alunos demoram para concluir o curso.



6. CONCLUSÃO

Se obteve durante esse um ano um excelente desenvolvimento do projeto. Os objetivos propostos ficaram bem claros, o caminho que se deveria seguir para atingi-los foi traçado e foi percorrido bem e sem muitos problemas.

As criações das bases de dados foram importantes e ocorreram da forma correta. Além do mais, as metas de elaboração das consultas foram realizadas e se obteve resultados importantes em cima disso. Pode-se fazer análises e tirar conclusões satisfatórias em cima dos resultados obtidos.

Por fim, no decorrer dos meses os avanços realizados colaboraram muito na criação do artigo. As pesquisas geraram retornos e alguns perfis de alunos ficaram claros, assim como seus comportamentos.

O artigo já está quase pronto, estando faltando apenas a inserção de alguns poucos gráficos que ainda estão sendo concluídos e a revisão do mesmo. Será questão de pouco tempo até que seja publicado.



REFERENCIAS BIBLIOGRÁFICAS

The First Online Portuguese Courses Launch on Coursera. Coursera Blog. 06 de maio de 2015. Disponível em: <<https://blog.coursera.org/the-first-online-portuguese-courses-launch-on/>>. Acesso em: 22 jul. 2018.

PAPPANO, Laura. **The Year of the MOOC.** The New York Times. 02 de nov. de 2012. Disponível em: <http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html?pagewanted=all&_r=0>. Acesso em: 22 jul. 2018.

DA MATTA, Cláudia Eliane. FIGUEIREDO, Ana Paula Silva. **MOOC: TRANSFORMAÇÃO DAS PRÁTICAS DE APRENDIZAGEM.** Publicações NEaD - Universidade Federal de Itajubá. 13 de junho de 2013. Disponível em: <https://nead.unifei.edu.br/images/conteudo/Artigos/ESUD_Claudia_AnaPaula.pdf>. Acesso em: 22 jul. 2018.

Andrew Ng, do Coursera: Como os MOOCs estão globalizando o conhecimento. Knowledge@Wharton. 15 de junho de 2015. Disponível em: <<http://www.knowledgeatwharton.com.br/article/andrew-ng-do-coursera-como-os-moocs-estao-globalizando-o-conhecimento/>>. Acesso em: 22 jul. 2018.

Conheça nossos parceiros. Coursera. Disponível em: <<https://www.coursera.org/about/partners>>. Acesso em: 22 jul. 2018.

CAPELAS, Bruno. **Coursera lança 24 cursos em português voltados ao mercado de trabalho.** Link - O Estado de S. Paulo. 27 de março de 2018. Disponível em: <<https://link.estadao.com.br/noticias/empresas,coursera-lanca-24-cursos-em-portugues-voltados-ao-mercado-de-trabalho,70002243559>>. Acesso em: 22 jul. 2018.