**Análise de dados do projeto olímpico**

### 1. Visão geral

Analisar dados históricos e atuais sobre as Olimpíadas Paraibana e Brasileira de Informática com o propósito de sumarizar seus resultados e caracterizar o perfil dos participantes campeões. Com a sumarização pretende-se entender a abrangência das olimpíadas no sentido do envolvimento de pessoas de diferentes cidades, escolas, culturas, gênero; qualidade em termos da média de pontos alcançados e premiações; impacto das olimpíadas na vida dos participantes sob a perspectiva de computação revelando quantos continuaram competindo, quantos entraram em computação@ufcg. Enquanto o perfil dos participantes tem como foco descobrir a motivação e as práticas dos medalhistas para assim repetir o exemplo e formar mais medalhistas.

### 2. Planejamento e Definição

#### 2.1 Plano de GQM

##### Objetivos/Metas/Questões/Métricas

* **Objetivo 1: Analisar** dados históricos e atuais sobre olimpíadas de informática **com o propósito de** sumarizar seus resultados **com respeito a** abrangência e competitividade das olimpíadas **do ponto de vista** de quem tem interesse nesse tipo de competição, em .particular, seus participantes **no contexto** das Olimpíadas de Informática realizadas na Paraíba (Olimpíada Paraibana de Informática e Olimpíadas Brasileira de Informática/Paraíba).
  + **Questão 1**: A quantidade de participantes das olimpíadas vem crescendo ao longo dos anos?
    - *Métricas*:
      * Número de inscritos por categoria e por ano.
  + **Questão 2**: Existem participantes das olimpíadas que ingressaram na computação@ufcg? Se sim, quais iniciativas ajudaram?
    - *Métricas*:
      * Número de competidores que ingressaram do curso de computação@ufcg
      * Lista de iniciativas que ajudaram a ingressar no curso
  + **Questão 3**: Os participantes treinados pelo projeto olímpico tendem a ganhar medalhas?
    - *Métricas*:
      * Valores da intersecção treinados pelo projeto olímpico / medalhistas
  + **Questão 4**: As olimpíadas reúnem participantes de diversas localidades, escolas e tipos de escolas (pública, privada, técnicas, não técnicas)?
    - *Métricas*:
      * Número de participantes classificados de acordo com a cidade de localização em ordem cronológica.
      * Número de participantes classificados por escola e tipo de escola (pública, privada, técnicas ou não técnicas) e em ordem cronológica.
  + **Questão 5:** Qual a qualidade das olimpíadas em termos de desempenho dos participantes ao longo dos anos? E existem diferenças entre o desempenho dos participantes recorrentes e os participantes novatos?
    - Métricas:
      * Desempenho médio (número de questões resolvidas) por categoria e por ano.
      * Desempenho individual, por categoria e por ano de novatos e recorrentes; correlação entre desempenho e recorrência na olimpíada
* **Objetivo 2: Analisar** dados de participantes campeões/medalhistas de olimpíadas de informática **com o propósito de** caracterizar o perfil desses medalhistas **com respeito** as suas motivações e práticas **sob o ponto de vista** de participantes dessas competições **no contexto** das Olimpíadas de Informática realizadas na Paraíba (Olimpíada Paraibana de Informática e Olimpíadas Brasileira de Informática/Paraíba).
  + **Questão 1**: Quais as estratégias de preparação mais comuns entre os participantes campeões recorrentes e os novatos?
    - *Métricas*:
      * Informações a respeito das práticas de preparação comuns entre os competidores que ganharam pela primeira vez um prêmio e os que já ganharam prêmios em outras olimpíadas de informática (por exemplo, material didático, ferramentas de apoio, horas de estudo, estuda sozinho ou em grupo, tem treinamento na escola ou não)
  + **Questão 2**: Qual o impacto dos treinamentos do projeto olímpico sobre o desempenho dos medalhistas?
    - *Métricas*:
      * Quantidades de questões respondidas dos medalhistas que participaram do projeto olímpico e dos que não participaram.
  + **Questão 3**: O incentivo mais comum para os medalhistas vem da escola?
    - *Métricas*:
      * Incentivos reais (por exemplo, treinamento na escola) ou incentivos simbólicos (por exemplo, estímulo moral dado pelos pais)
  + **Questão 4**: Existe um perfil que caracteriza os medalhistas antes de começarem a participar de olimpíadas de informática?
    - *Métricas*:
      * Aspectos acadêmicos: por exemplo, disciplinas que mais gostam na escola, desempenho escolar, participa de outras olimpíadas de conhecimento;
      * Aspectos sociais: por exemplo, como ele faz amigos, se pratica esportes, se participa de projetos sociais.
  + **Questão 5**: Qual a relação dos medalhistas com a computação?
    - *Métricas*:
      * Indicadores de uso da computação: programador, uso de redes sociais e internet, ...
  + **Questão 6**: Quais as motivações para os medalhistas continuarem competindo em olimpíadas de informática?
    - *Métricas*: a lista das motivações sejam pessoais, profissionais ou acadêmicas.

#### 2.2 Plano de Medição

Com base nas métricas definidas anteriormente, será necessário coletar/recuperar dados:

* **Disponibilidade dos dados?**
  + queremos coletar mais dados de 2016?

**Dados históricos e atual das olimpíadas OPI a partir de 2010**

* **Tipos de dados**
  + dados pessoais dos participantes (parcial e fragmentado): nome, cidade, escola de origem, gênero, serie;
  + dados sobre a participação de cada um nas competições (parcial e fragmentado): classificação (posição no ranking) e desempenho (quantidade de pontos acumulados.
* **Quando**
  + os dados de 2010 a 2015 já foram coletados;
  + os dados de 2016 serão coletados no ato da inscrição dos participantes nas competições e a partir dos resultados divulgados da competição.
    - maio2016 (Fase 1 do ensino fundamental e médio)
    - ago2016 (ensino superior)
    - set2016 (Fase 2 do ensino fundamental e médio)
* **Como serão coletados (instrumentos de medição)**
  + serão coletados via formulário de inscrição para cada competição e no site oficial da OPI
  + via documentos divulgados com os resultados dos participantes nas competições
  + obs.: verificar com Rohit onde estão armazenados e recuperar
* **Por quem**
  + inscrições dos participantes são realizadas de formas distintas de acordo com a categoria pelas suas escolas de origem, via formulário online.
  + resultados das competições: Gleyser e Hugo

**Dados históricos e atual das olimpíadas OBI na Paraíba a partir de 2000**

* **Tipos de dados**
  + dados pessoais dos participantes (somente 2016 e 2015): nome, cidade, escola de origem, gênero, serie;
  + dados sobre a participação de cada um nas competições: classificação (posição no ranking) e desempenho (quantidade de pontos acumulados.
* **Quando**
  + os dados de 2000 a 2014 já foram coletados e temos somente dos medalhistas
  + os dados de 2015 já foram coletados e temos mais detalhes
  + os dados de 2016 serão coletados no ato da inscrição dos participantes nas competições e a partir dos resultados divulgados da competição.
    - maio2016 (Fase 1 do ensino fundamental e médio)
    - ago2016 (ensino superior)
    - set2016 (Fase 2 do ensino fundamental e médio)
* **Como serão coletados (instrumentos de medição)**
  + serão coletados via formulário de inscrição para cada competição e no site oficial da OBI
  + via documentos divulgados com os resultados dos participantes nas competições
  + obs.: verificar com Rohit onde estão armazenados e recuperar
* **Por quem**
  + inscrições dos participantes são realizadas de formas distintas de acordo com a categoria pelas suas escolas de origem, via formulário online.
  + resultados das competições: Gleyser e Hugo
* **Se relaciona com quais métricas:** O1Q1, O1Q2, O1Q5

**Dados históricos e atuais dos alunos da computacao@UFCG a partir de 2009**

* **Tipos de dados:** nome, ano/período de ingresso no curso
* **Quando:**
  + existe uma possibilidade de reusar os dados da PRE (já coletados); precisa de autorização
  + dados a mais a definir
* **Como serão coletados:**
  + Verificando com o grupo Analytics a listagem dos alunos que entraram no curso a partir de 2001 (por conta da OBI) e fazendo uma referência cruzada com os participantes de olimpíadas desde 2008
  + Por meio de entrevistas para descobrir as iniciativas que colaboraram para os alunos ingressarem no curso
* **Por quem:** Gleyser e Hugo
* **Se relaciona com quais métricas:** O1Q2

**Dados históricos e atual sobre os treinamentos do projeto olímpico oferecidos desde 2015**

* **Tipos de dados**
  + nome, qual treinamento/curso, data de realização, desempenho no curso (parcial; somente para os treinamentos do pet)
* **Quando**
  + os dados de 2015 já foram coletados
  + os dados de 2016 serão coletados a medida que os cursos acontecem:
    - semana do fera 2016.1 e 2016.2
    - treinamentos permanentes nas escolas a partir de 2015
    - sob demanda a partir de 2015
* **Como serão coletados**
  + Os dados serão coletados dos relatórios dos cursos e das ferramentas utilizadas e fazendo uma referência cruzada com os participantes de olimpíadas em 2015 e 2016
* **Por quem**
  + Gleyser e Hugo
* **Se relaciona com quais métricas:** O1Q2, O2Q2

**Dados dos medalhistas de 2015 e 2016**

* **Tipos de dados**
  + material usado, tempo de estudo, treinamento sozinho ou na escola, incentivos, aspectos acadêmicos e socais, indicadores de uso da computação, motivações para competir, etc.
* **Quando**
  + os dados de 2015 já foram coletados
  + os dados de 2016 serão coletados na cerimônia (novembro); e até outubro por meio das entrevistas com os medalhistas
* **Como serão coletados**
  + Recuperar os dados da pesquisa realizada em 2015 e replicar o estudo em 2016, porém além do survey, usar entrevistas estruturadas (com medalhistas de maior destaque).
* **Por quem:** 
  + Gleyser e Hugo
* **Se relaciona com quais métricas:** O2Q1, O2Q3,O2Q4 e O2Q5

#### 2.3 Plano de Análise

**Como analisar os dados (quais as respostas que gostaríamos de obter)?**

Será usada a estratégia empírica por meio da observação e análise dos dados. Os dados serão analisados usando a estatística descritiva. Serão realizados procedimentos estatísticos para eliminação de outliers, análise de distribuições e aplicação da inferência estatística (teste-t, ANOVA, regressão).

Iremos pensar nas respostas que gostaríamos de obter… (o que esperar?)

**Como agregar os dados?**

Como apresentar os dados (gráficos, tabelas, …)

Gráficos criados seguindo os princípios da integridade gráfica...

### 3. Coleta de Dados

### 4. Análise e Interpretação