



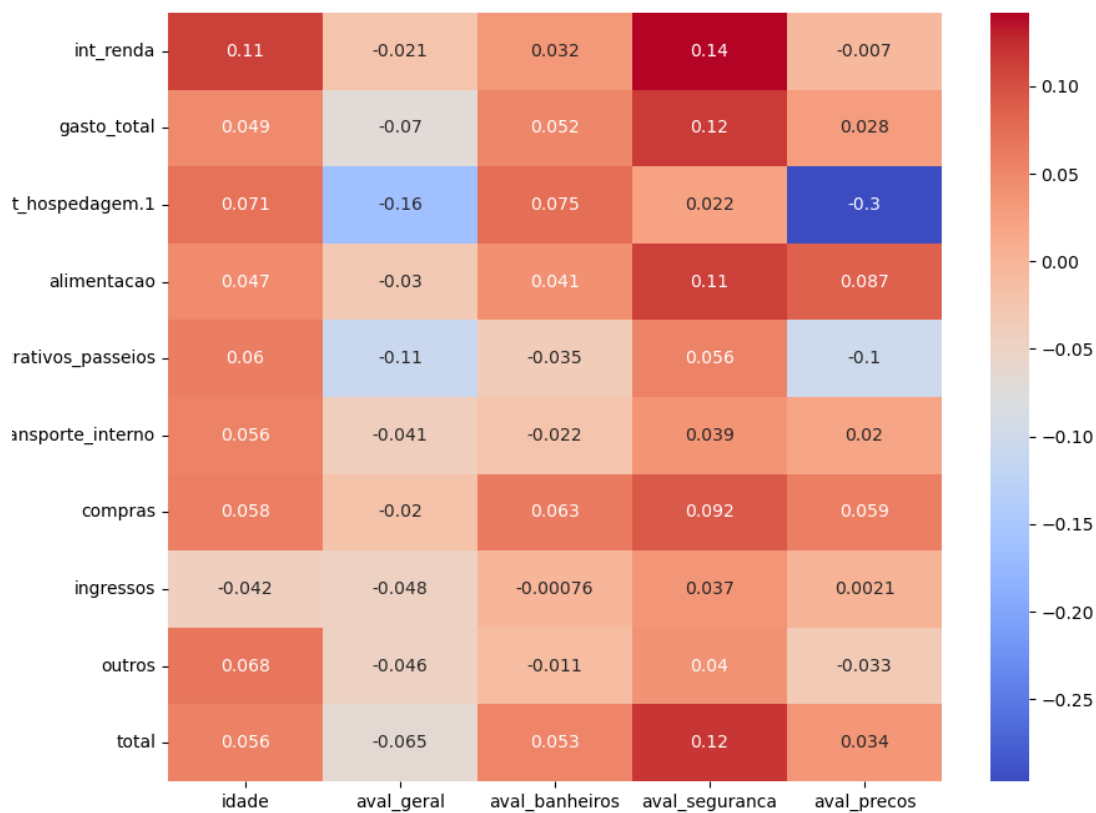
Relatório

O Carnaval Pudding foi baseado em um conjunto de dados com informações sobre a participação no carnaval de Belo Horizonte em 2018. As informações incluem características demográficas dos participantes, como idade, país, estado e cidade de origem, assim como informações sobre os gastos com a participação no carnaval, como gastos com hospedagem, alimentação e transporte interno. O conjunto de dados também contém informações sobre a opinião dos participantes em relação ao carnaval, incluindo avaliações de segurança, preços e banheiros disponíveis, bem como se os participantes pretendem participar novamente no futuro e o que eles acharam do carnaval em geral. As informações foram coletadas por meio de questionários e entrevistas com os participantes do carnaval.

Com base em um conjunto de dados carregado em um dataframe, transformado e tratado para remover dados nulos e outliers, o grupo de ciência de dados conseguiu identificar padrões interessantes sobre o comportamento dos usuários do site em relação ao carnaval. Utilizando técnicas de clusterização e machine learning, foi possível criar um modelo que aponta o bloco do usuário com base em sua faixa de renda, estado civil, idade e gênero.

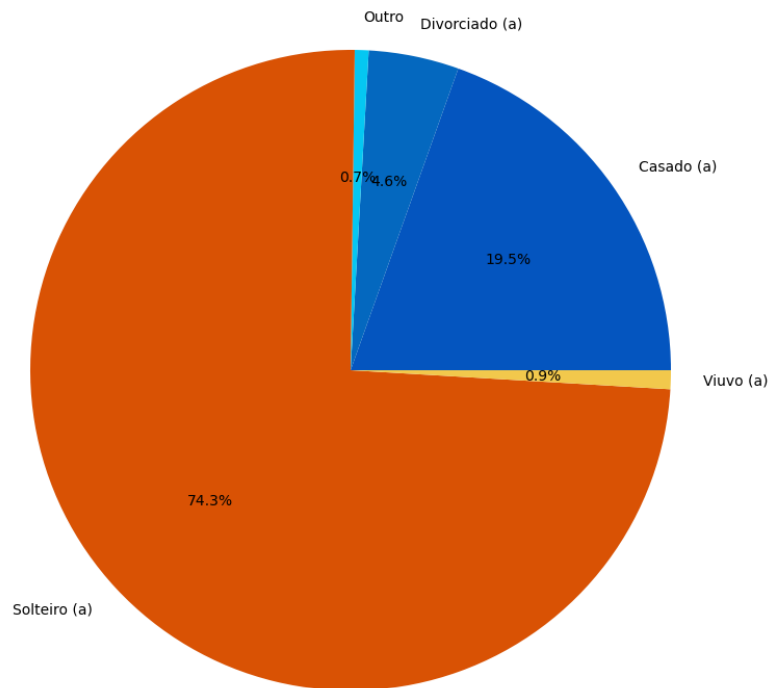
Entre as principais descobertas do estudo, destaca-se que o carnaval mantém uma avaliação geral parecida para todos os tipos de pessoa, o que sugere que o evento é amplamente apreciado por diversos grupos. Além disso, foram identificados alguns estereótipos, chamados de "figurinhas carimbadas", que tendem a se destacar entre os usuários do site.

Outra descoberta interessante é que pessoas que gastam menos tendem a gostar mais ainda do carnaval, o que pode indicar que o evento é visto como uma oportunidade de diversão acessível para esses grupos. Por fim, o estudo revelou que pessoas com mais de 65 anos também tendem a apreciar o carnaval, desmentindo a ideia de que essa faixa etária não tem interesse pelo evento.

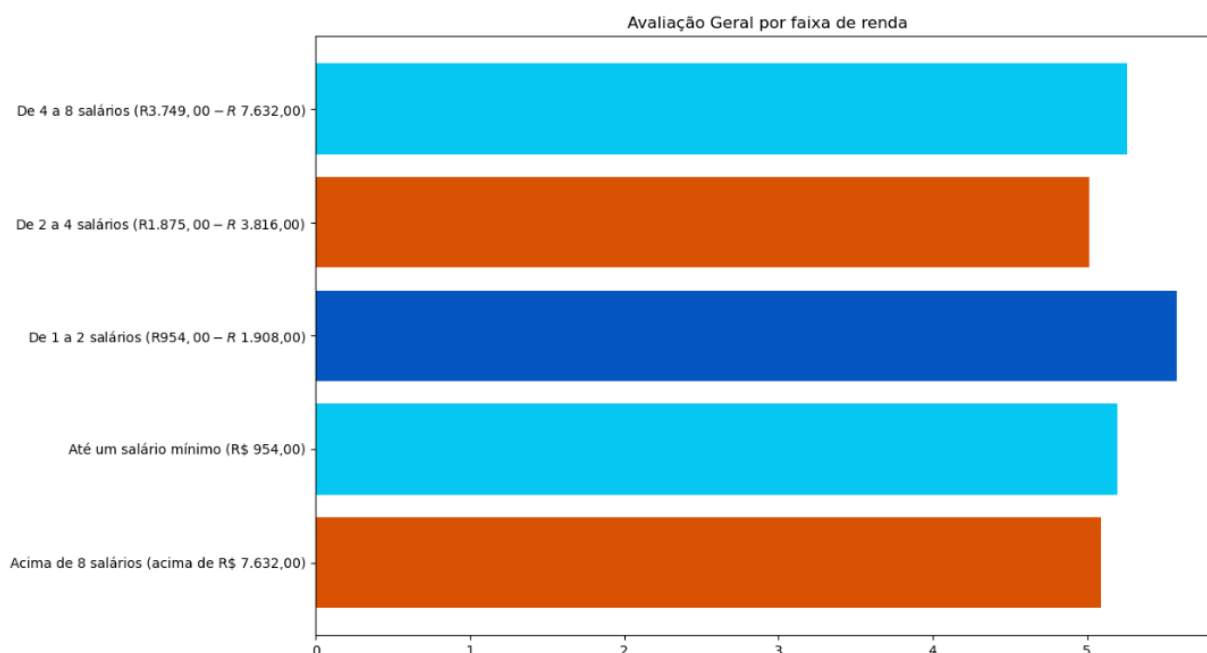


Foi realizado um gráfico de correlação na AED para comprovar os fatos citados, e a maior correlação foi do preço de hospedagem e a baixa na avaliação dos preços da viagem. Então foi citada uma dica no site para procurar um camarada que possa oferecer hospedaria. ("Se arranjar uma hospedagem na casa de alguém fica bem mais em conta!")

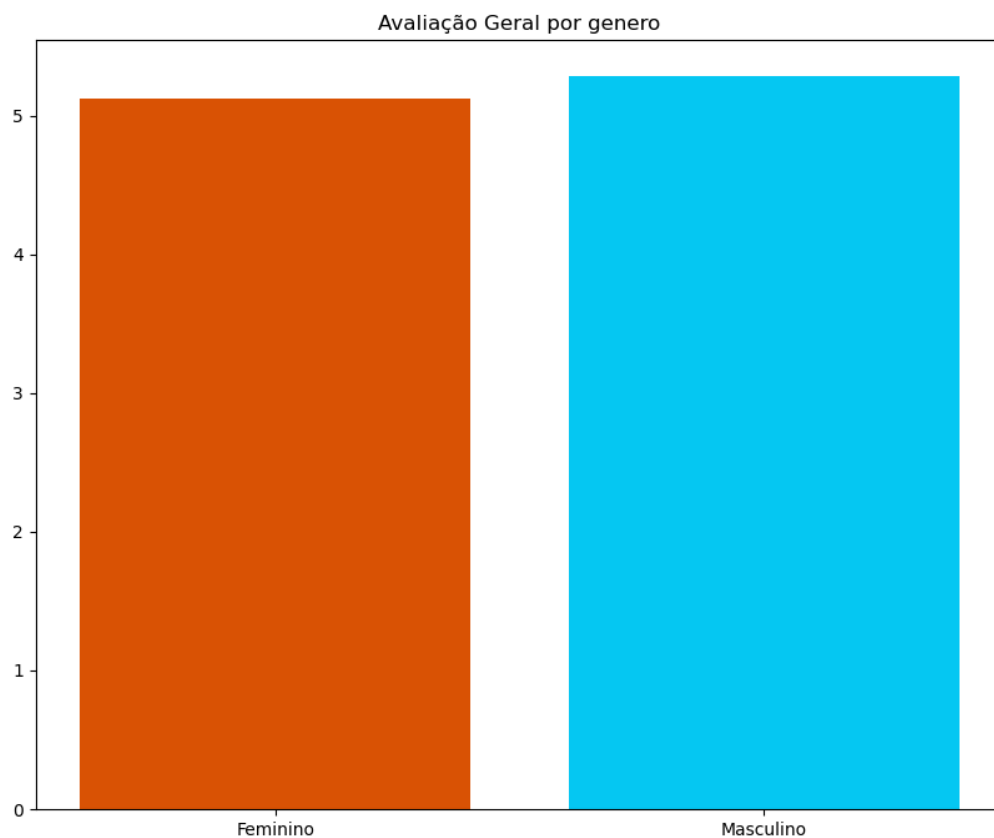
Percentual de frequentadores de cada estado civil



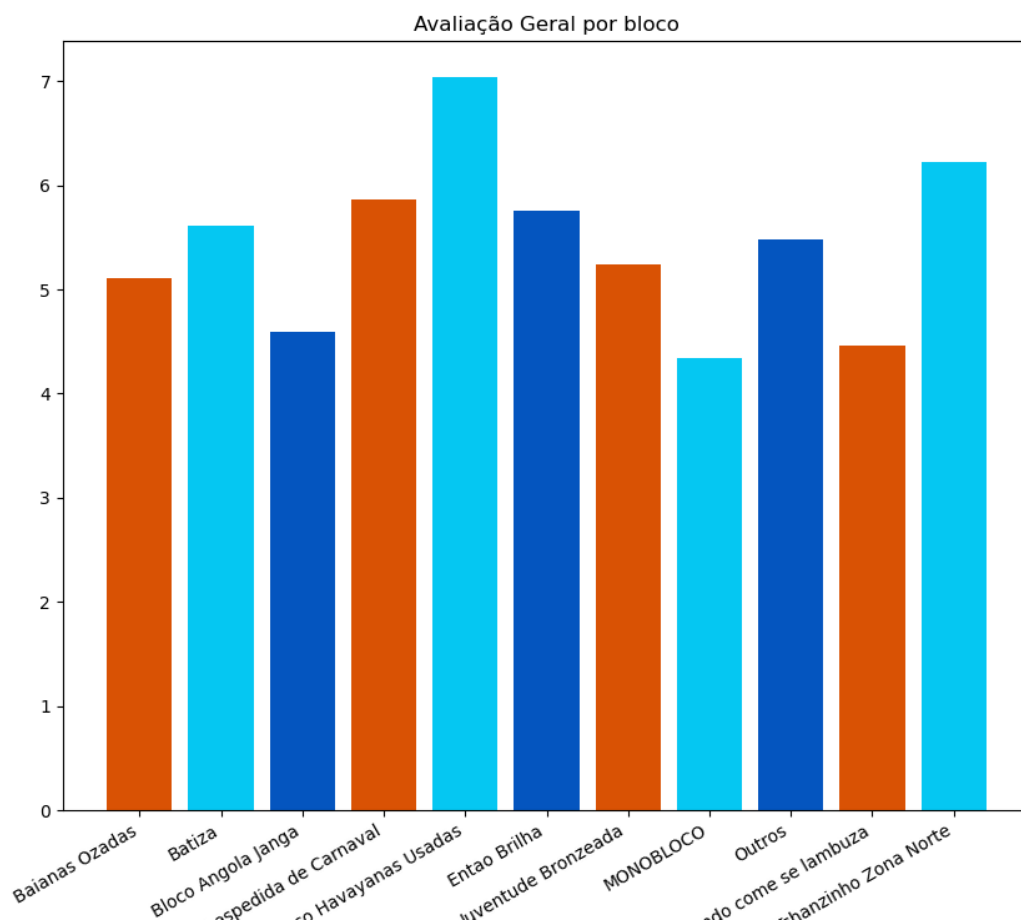
Agora para valores qualitativos, escolhemos mostrar graficamente que todas as categorias estão significativamente presentes. O gráfico de pizza mostra que apesar de solteiros serem a maioria no carnaval, também estão presentes de Divorciado a Viúvo, e até “outros”. Prosseguindo com a ideia de que o carnaval é democrático com qualquer um, e também uma oportunidade para os casais se divertirem.



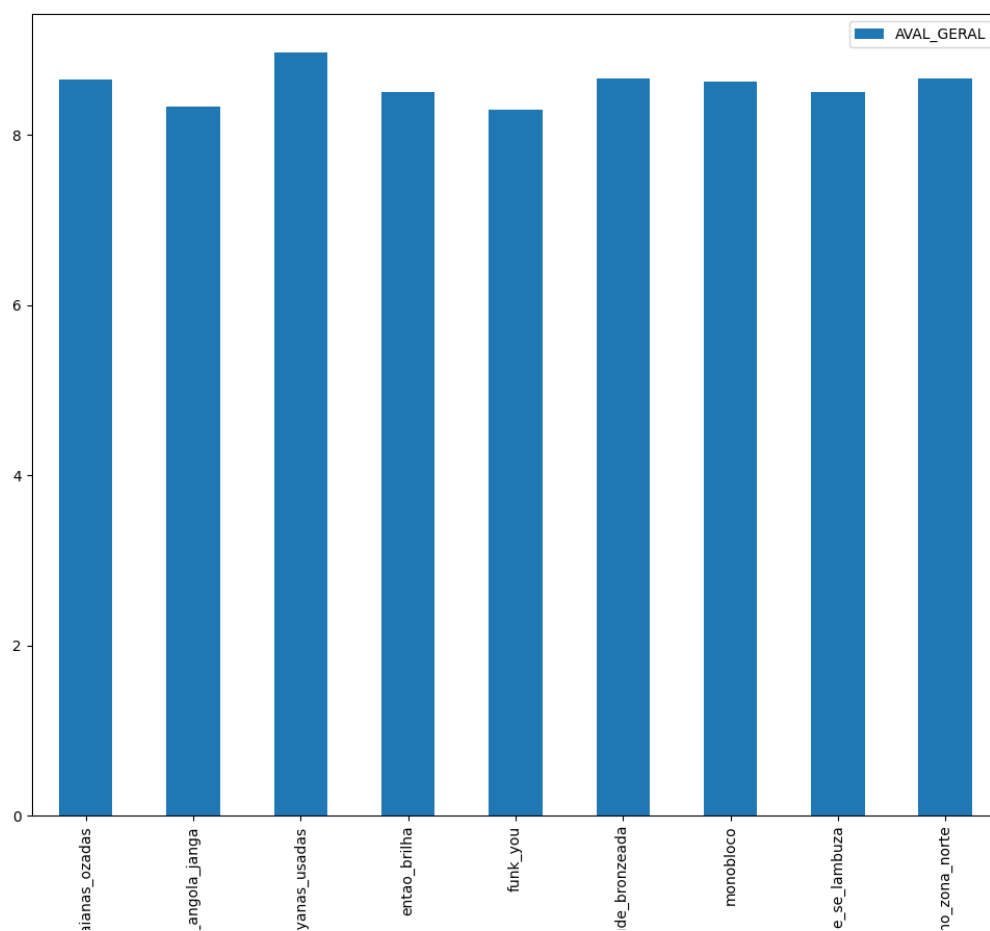
O estudo sugere que o carnaval é amplamente apreciado por diversos grupos, e que a faixa de renda não tem influência nas avaliações, o que indica que o evento não é elitista nem exclusivamente popular. Essa descoberta é importante, pois mostra que a festa é acessível a pessoas de diferentes classes sociais.



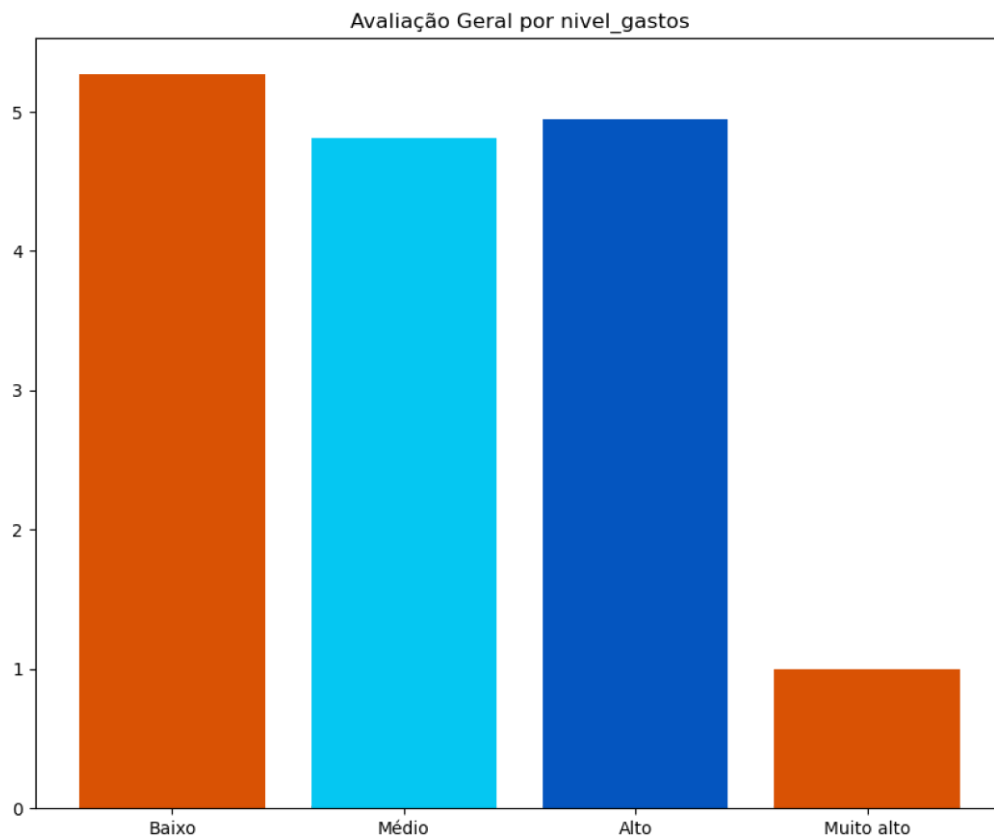
De acordo com o gráfico de colunas, pode se concluir que as pessoas avaliaram os aspectos do desfile de forma similar, independente do gênero. Isso porque, como mencionado anteriormente, as médias das avaliações de outras categorias apresentaram diferenças pouco expressivas entre os gêneros feminino e masculino.



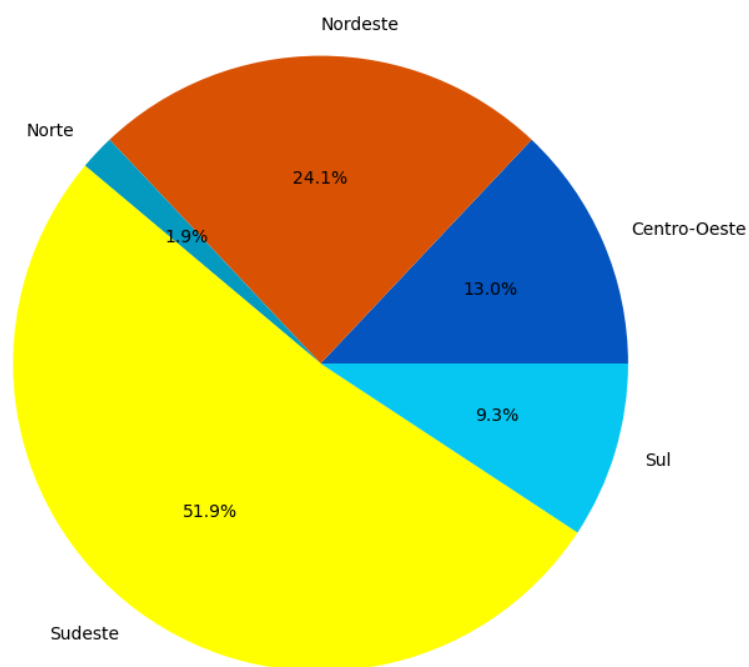
Compreender o comportamento das pessoas em relação aos blocos de carnaval é fundamental para oferecer uma experiência mais personalizada e satisfatória para os usuários do site. No entanto, é preciso ter em mente que a avaliação dos blocos pode variar bastante de um ano para o outro. Isso significa que as avaliações não são suficientemente confiáveis para serem a única fonte de informação na hora de indicar ou deixar de indicar algum bloco para o usuário.



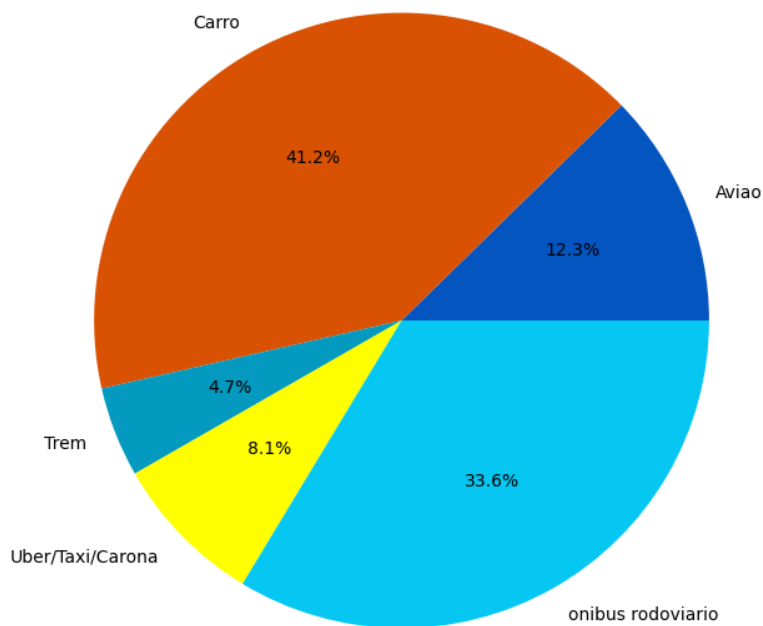
Existem diversos fatores que podem influenciar na avaliação de um bloco, como as condições climáticas, a qualidade das atrações, a quantidade de pessoas presentes, dentre outros. Além disso, a percepção das pessoas em relação a um mesmo evento pode mudar ao longo do tempo, o que pode refletir em avaliações diferentes em anos diferentes, como a demonstrada nesse gráfico.



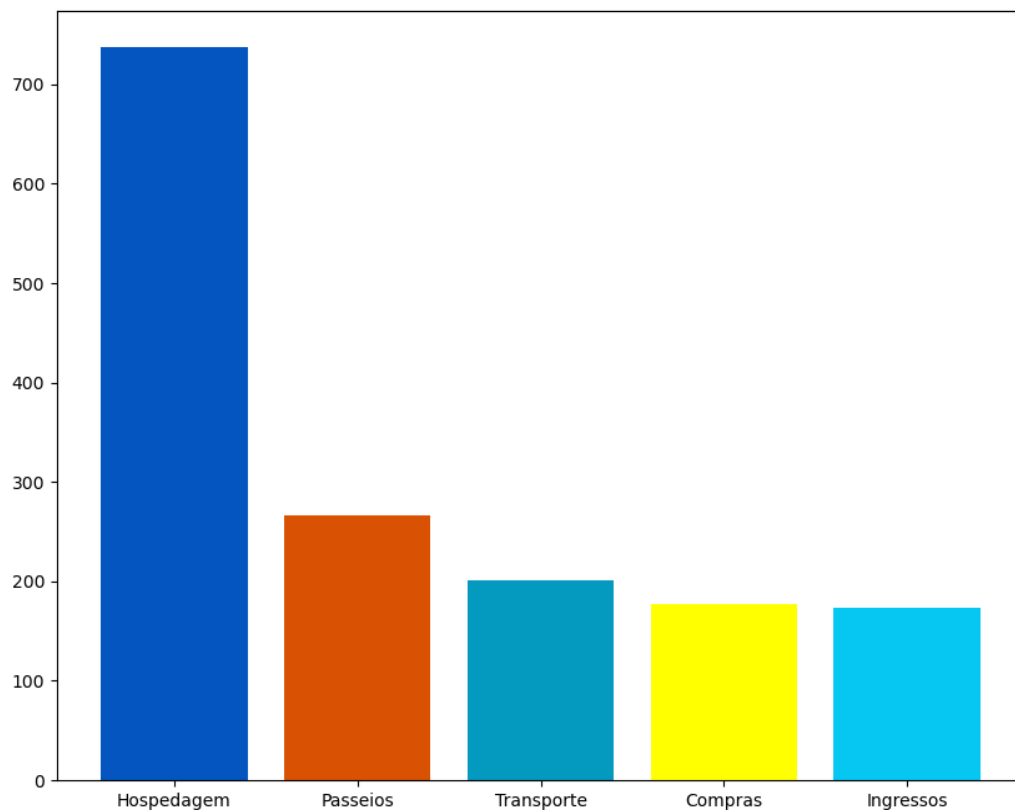
O gráfico apresenta a distribuição das avaliações do Carnaval de Belo Horizonte por faixa etária. É possível observar que, em geral, as avaliações são bastante parecidas entre as diferentes faixas de idade, indicando que o evento é apreciado por pessoas de todas as idades. No entanto, nota-se que as pessoas com mais de 65 anos avaliaram o carnaval ligeiramente melhor do que as outras faixas etárias, o que pode indicar que essas pessoas têm uma apreciação particular pela festa ou que a experiência do carnaval é mais significativa para elas. As informações apresentadas no gráfico são importantes para entender como diferentes grupos etários avaliam o carnaval e podem ajudar na elaboração de estratégias para tornar o evento mais inclusivo e atraente para pessoas de todas as idades.



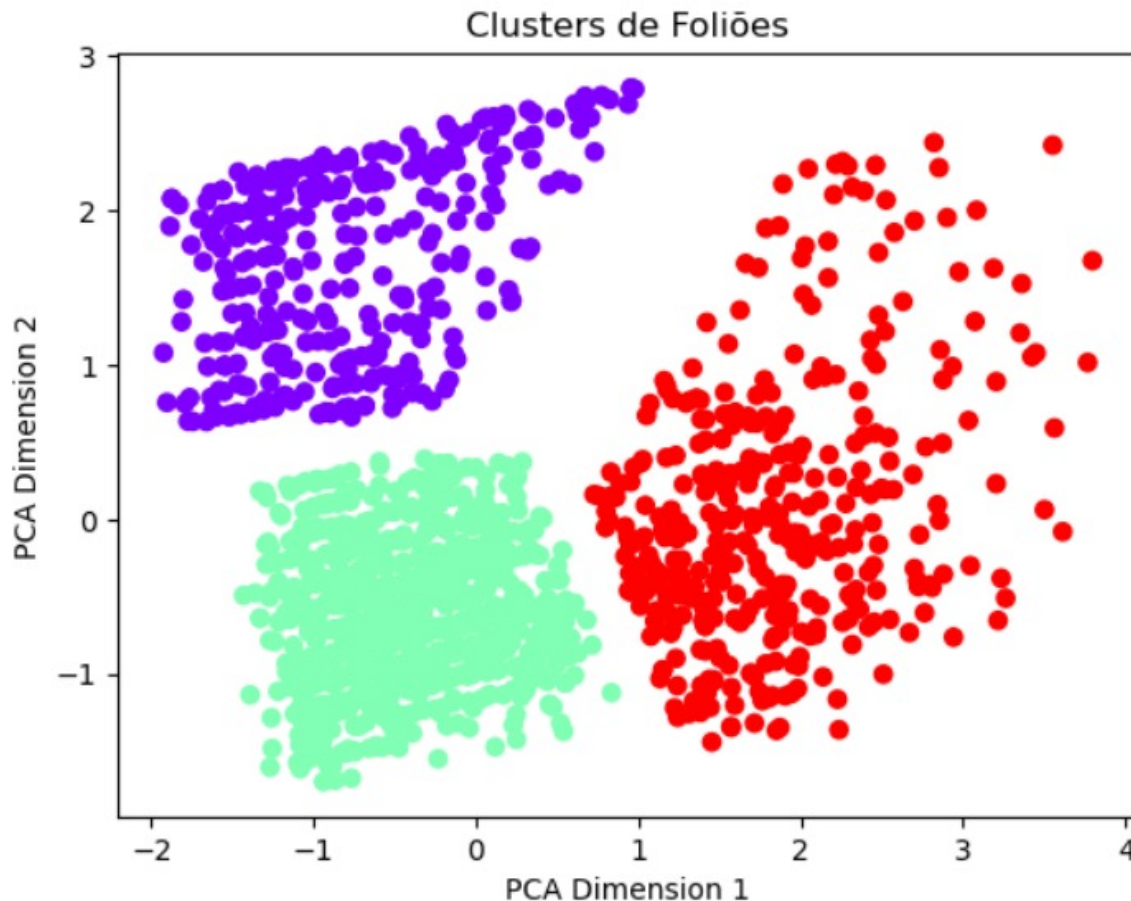
O gráfico de pizza apresentado no relatório mostra que a maioria dos participantes do Carnaval de Belo Horizonte é do Sudeste do Brasil. No entanto, é importante notar que pessoas de todas as regiões do país também participam do evento, como é possível observar nas fatias correspondentes às regiões Norte, Nordeste, Centro-Oeste e Sul.



O gráfico apresenta os meios de transporte mais utilizados pelas pessoas que chegaram em Belo Horizonte por meio de carona. Os dados mostram que a maioria das pessoas utilizou o carro como meio de transporte, seguido por ônibus rodoviário, avião, Uber/Táxi/Carona e, por último, o trem. Isso pode indicar que as pessoas preferem a conveniência e flexibilidade do carro ou ônibus rodoviário em relação às opções mais caras, como avião e táxi. O trem, por outro lado, parece ser uma opção menos popular provavelmente devido à limitação de rotas e horários disponíveis, mas adicionamos a rota como curiosidade no site.



Analisando o gráfico de barras, é possível observar que os gastos com hospedagem são um dos principais fatores que influenciam no orçamento de uma viagem. No entanto, é importante ressaltar que é possível economizar consideravelmente nesse aspecto ao optar por se hospedar na casa de amigos ou parentes. É uma alternativa que pode reduzir significativamente os gastos com hospedagem e permitir que mais recursos sejam direcionados para outras atividades durante a viagem. Portanto, é fundamental começar a planejar a viagem com antecedência, considerando todas as possibilidades disponíveis para economizar e aproveitar ao máximo a experiência.



O gráfico de clusterização apresenta uma divisão clara de três perfis distintos gerados por modelo de clusterização não supervisionado. Primeiro, o número de clusters a usar foi determinado pelo *elbow method* ("método de cotovelo") e a clusterização feita por k-means. Para apresentação gráfica da divisão foi efetuada uma *principal component analysis* de 2 dimensões para apresentar o gráfico acima. Depois, para determinar quais características são mais preponderantes na separação dos clusters, foi usada a técnica de Machine Learning de Regressão Logística para encontrar os pesos de cada quocientes. Cada perfil apresenta características e comportamentos bem definidos, permitindo uma análise precisa e identificação de padrões em cada um. Essa ferramenta pode ser extremamente útil em diversas áreas, desde o marketing para entender o perfil dos consumidores, até a área médica para identificar grupos de pacientes com características semelhantes e, assim, desenvolver tratamentos mais eficazes. A clusterização se torna uma poderosa aliada na tomada de decisões, permitindo uma abordagem mais direcionada e assertiva em cada caso específico.

Conclusão

A análise dos dados do Carnaval de BH em 2018 permitiu uma visão mais clara dos principais aspectos desse evento, como os meios de transporte utilizados pelos foliões, as principais fontes de gastos e as avaliações dos festeiros. Porém, ainda há muito a explorar nesses dados, a equipe almejava obter mais tempo para analisar fontes de hipóteses como a geolocalização. Por exemplo, seria possível avaliar se a presença de um bloco influencia a participação de outros, ou se há potencial para aumentar a participação em outras áreas da cidade. Essas informações são cruciais para o planejamento de eventos futuros e podem auxiliar na melhoria da experiência dos foliões.

Obtivemos dados do carnaval de bh de 2019 pela mesma fonte do dataset fornecido, o grupo com mais tempo realizaria mais análises comparando a evolução do carnaval. Entendo quais são suas tendências no decorrer do tempo e como apontar padrões para a melhoria de experiência do folião.

Em suma, o estudo sobre o comportamento dos participantes do Carnaval de Belo Horizonte em 2018, utilizando técnicas de ciência de dados, mostrou que o evento é amplamente apreciado por diversos grupos, independentemente da faixa de renda, gênero ou estado civil. As descobertas também revelaram alguns estereótipos entre os participantes e indicaram que o carnaval é visto como uma oportunidade de diversão acessível para aqueles que gastam menos. Além disso, o estudo desmentiu a ideia de que pessoas com mais de 65 anos não têm interesse pelo evento. A partir dessas informações, é possível oferecer uma experiência mais personalizada e satisfatória para os usuários do site, apresentando dados pertinentes e uma interatividade associadas com tecnologias confiáveis de predição para proporcionar a melhor experiência para o usuário.