



GIA Rede Gazeta



Controle do Documento

Histórico de revisões

Data	Autor	Versão	Resumo da atividade
09/08/2022	João Marques e Raphael Antunes	0.1	Canvas de proposta de valor
04/08/2022	Raduan Muarrek e Luana Parra	0.1.2	Análise de indústria
04/08/2022	Vitor Oliveira e João Suarez	0.1.3	Matriz de riscos
08/08/2022	João Marques, Raphael Antunes, Luana Parra	0.1.4	Matriz SWOT
09/08/2022	João Marques e Raphael Antunes	0.1.5	Canvas de proposta de valor
10/08/2022	Raduan Muarrek e João Marques	0.1.6	Descrição da predição
10/08/2022	Luana Parra e Vitor Oliveira	0.1.7	Descrição dos dados
11/08/2022	Elisa Flemer e Raphael Antunes	0.1.8	Criação dos gráficos para a análise de dados
11/08/2022	Luana Parra	0.1.9	Revisão dos artefatos da Sprint 1
12/08/2022	Elisa Flemer	0.2	Revisão final para Sprint 1

Sumário

1. Introdução	5
2. Objetivos e Justificativa	6
2.1. Objetivos	6
2.2. Justificativa	6
3. Metodologia	7
3.1. CRISP-DM	7
3.2. Ferramentas	7
3.3. Principais técnicas empregadas	7
4. Desenvolvimento e Resultados	8
4.1. Compreensão do Problema	8
4.1.1. Contexto da indústria	8
4.1.2. Análise SWOT	8
4.1.3. Planejamento Geral da Solução	8
4.1.4. Value Proposition Canvas	8
4.1.5. Matriz de Riscos	8
4.1.6. Personas	9
4.1.7. Jornadas do Usuário	9
4.2. Compreensão dos Dados	10
4.3. Preparação dos Dados	11
4.4. Modelagem	12
4.5. Avaliação	13
4.6. Comparação de Modelos	14
5. Conclusões e Recomendações	14
6. Referências	15
Anexos	16

1. Introdução

A Rede Gazeta de Comunicações é a maior empresa de comunicação do Espírito Santo, com mais de 500 funcionários. Fundada em 1928 com o jornal *A Gazeta*, tem como maior propósito informar, entreter e prestar serviços de comunicação aos capixabas com qualidade, ética e inovação, contribuindo para o desenvolvimento socioeconômico, cultural e de cidadania. Atualmente, o grupo é formado por um site de notícias (www.agazeta.com.br); oito rádios; quatro emissoras de TV aberta (TV Gazeta) afiliadas à Rede Globo; e dois portais de notícias locais (G1 Espírito Santo e o Globo Esporte Espírito Santo).

No que tange às emissoras, devido à sua conexão com a Rede Globo, devem coordenar a programação local com os eventos nacionais produzidos pela rede. Assim, com slots limitados para veicular seus próprios conteúdos, torna-se imprescindível prever a audiência de novos eventos com acurácia para maximizar o retorno de seus investimentos. Objetiva-se, com isso, escolher sempre o melhor evento para cada horário disponível e alocar recursos eficientemente, priorizando campanhas de marketing para programas com menor audiência esperada.

No momento, apesar de existirem registros históricos de score de audiência para diferentes horários e demográficos, não há análises de correlação entre programa (com suas características principais) e slot. Consequentemente, faltam evidências para prever e justificar a transmissão de certos eventos em certos horários, causando significativa incerteza para os stakeholders da empresa a cada modificação na grade de programação.

2. Objetivos e Justificativa

2.1. Objetivos

O principal objetivo da Rede Gazeta é maximizar o retorno financeiro de cada evento veiculado. Essa monetização se dá através da venda de slots publicitários, cuja precificação, por sua vez, é fortemente influenciada do score de audiência do programa sendo veiculado durante o slot comercial. Assim, em termos mais específicos, para maximizar seus retornos, a Rede Gazeta deve maximizar a audiência de cada conteúdo a fim de atrair mais anunciantes.

Para atingir esses objetivos, convém também saber quais variáveis mais influenciam as audiências de cada horário com o fito de priorizá-las na criação de novos programas e suprir as expectativas do público. Além disso, faz-se necessário um entendimento profundo das audiências esperadas para novos eventos para se alocar recursos de marketing de modo a atrair mais telespectadores àqueles com menor expectativa.

2.2. Justificativa

A nossa proposta de solução é a GIA (Inteligência Artificial da Rede Gazeta), um modelo preditivo o qual irá prever a audiência de programas futuros, além de ser uma ferramenta para medir o esforço necessário para determinadas programações. Ademais, com o auxílio das tabelas fornecidas pela Rede Gazeta é possível identificar as variáveis que impactam a audiência. Portanto, o resultado final é fornecer um score de audiência e o que contribuiu para esse resultado, a partir das variáveis, data, horário, alcance e gênero do programa.

3. Metodologia

Descreva as etapas metodológicas que foram utilizadas para o desenvolvimento, citando o referencial teórico. Você deve apenas enunciar os métodos, sem dizer ainda como ele foi aplicado e quais resultados obtidos.

3.1. CRISP-DM

Descreva brevemente a metodologia CRISP-DM e suas etapas de processo

O CRISP-DM (Cross Industry Standard Process for Data Mining) é uma metodologia ágil voltada para projetos envolvendo Machine Learning, mineração e análise de dados. Sendo assim, é um processo cíclico dividido nas seguintes etapas:

1. "Business Understanding" (estudo do projeto ou negócio, atendendo os objetivos e interesses do cliente);

2."Data Understanding (identificar, coletar e analisar os conjuntos de dados que podem ajudar a atingir os objetivos do projeto);

3."Data Preparation" (ocorre a manipulação de dados, filtrando quais dados serão usados para a modelagem);

4."Modeling" (desenvolver um modelo e selecionar a técnica de modelagem);

5."Evaluation" (avaliar a qualidade, fidedignidade e segurança dos resultados obtidos da etapa de Modelagem);

6."Deployment" (se inicia o processo de desenvolvimento dos modelos criados e avaliados nas etapas anteriores)

3.2. Ferramentas

Descreva brevemente as ferramentas utilizadas e seus papéis (Google Collaboratory)

Nome	O que é	Em que foi utilizado	Versão
Google Collaboratory	É um serviço de nuvem gratuito hospedado pelo próprio Google para incentivar a pesquisa de Aprendizado de Máquina e Inteligência Artificial.		

3.3. Principais técnicas empregadas

Descreva brevemente as principais técnicas empregadas, algoritmos e seus benefícios

4. Desenvolvimento e Resultados

4.1. Compreensão do Problema

4.1.1. Contexto da indústria

A indústria televisiva aberta opera, primordialmente, em modelo B2B através da venda de slots publicitários para anunciantes diversos nos intervalos de cada evento (publicidades, merchandising, comerciais). A receita dessa atividade é diretamente relacionada ao score de audiência atingido pelos programas veiculados pela emissora. Nesse sentido, a atenção dos telespectadores é a mercadoria ofertada a seus clientes, isto é, às empresas que buscam divulgar seus produtos e serviços para um público de larga escala.

Dentre os principais players do mercado, tem a Globo, RecordTV, SBT, Band e TV Brasil. Destes, a Globo apresenta as maiores taxas de audiência em quase todas as situações, beirando um monopólio da indústria. A lista abaixo apresenta mais detalhes sobre cada uma delas.

Globo: Assistida por mais de 200 milhões de pessoas diariamente (no Brasil e no exterior), é a segunda maior rede de televisão comercial do mundo cobrindo 99,55% do total da população brasileira. Tem como diferenciais o jornalismo (especialmente o Jornal Nacional e Fantástico) e a produção de telenovelas. Possui 120 afiliadas e inúmeros canais, incluindo plataformas de streaming e transmissão para outros países.

Rede Record: Fundada em 1993 pelo empresário Paulo Machado de Carvalho, a Rede Record é uma rede de televisão aberta que passa em canal aberto e tem como nicho conteúdo religioso. Está na segunda colocação de audiência nacional pelo IBOPE e possui um trabalho de digitalização bem avançado, já tendo uma plataforma de streaming e uma grande presença nas redes sociais.

SBT: O SBT foi fundado pelo empresário e comunicador Sílvio Santos em 1981 após o Grupo Sílvio Santos ganhar uma licitação do Governo Federal para comprar a então “TV Tupi”. Hoje, ela ocupa a terceira posição na média da audiência nacional medida pelo IBOPE. Podemos considerar que o SBT ocupa dois nichos: o de programa de auditório, como programa do Ratinho e o próprio programa do Sílvio Santos, e o de programas infantis, como a novela *Chiquititas* e desenhos animados.

Rede Bandeirantes: A rede Bandeirantes foi criada em 1967 por João Jorge Saad após a compra da Rádio Bandeirantes. Hoje, a quarta posição na média nacional do IBOPE. Atua historicamente com esportes em geral e também com política, sendo tradição a produção dos primeiros debates de cada temporada de eleição no Brasil.

TV Brasil: É a emissora aberta oficial do Poder Executivo Brasileiro, além de ser obrigatoriamente transmitida em operadoras de TV paga. Atualmente, televisa uma mistura de programas independentes, seriados e transmissão de eventos esportivos.

No que tange às tendências na indústria, a maior delas é a transformação digital. Com o advento de plataformas de streaming, a indústria televisiva tem se reinventado para disponibilizar sua grade na internet ao vivo e sob demanda. Um exemplo é a Globo Play, que oferece não só produções originais como também filmes e séries internacionais para competir com gigantes tais quais Netflix e Amazon Prime.

Além disso, tem-se o crescimento significativo da personalização no consumo de conteúdo. Através de algoritmos de predição, os principais streamings e redes sociais já conseguem recomendar o conteúdo mais relevante para cada usuário de modo certo. Assim, há grande pressão para que a indústria televisiva tradicional também agregue mais personalização em suas operações.

Ademais, o Brasil hoje oferece solo fértil para a produção de séries ficcionais e não ficcionais. Seguindo o sucesso desse modelo em outros países, existe significativa motivação para investir em documentários e minisséries nacionais, como demonstrado pelo sucesso de títulos tais quais “Irmandade” e “3%”.

Análise aprofundada seguindo o modelo de 5 forças de Porter

Ameaça de novos entrantes: O estabelecimento de uma nova emissora de televisão exige grande capital inicial – com 70% de acionistas brasileiros –, uma concessão da Anatel e uma aprovação do Ministério de Comunicações quanto à sua proposta de programação e condições técnicas e financeiras. Sem dúvida, é uma empreitada que exige muitos recursos, paciência e competência. Além disso, uma vez estabelecida, uma nova emissora teria ainda o desafio de concorrer com as grandes redes de televisão brasileiras por audiência. Nesse sentido, percebe-se que há barreiras significativas para novos entrantes, de modo que a ameaça destes é baixa. No mercado atual existem grandes players, o que cria uma barreira de novos entrantes com alta capacidade de crescimento. Desse modo, por mais que hoje a produção de conteúdo seja de fácil acesso, a distribuição (caso da rede gazeta) é um espaço com poucas empresas e que possuem quase um monopólio no segmento. Por fim, a possibilidade de novos entrantes é baixa, pois existe uma demanda de capital inicial, caixa e competência de desafiar grandes competidores nesse mercado.

Poder de barganha dos fornecedores: Os fornecedores são os produtores de conteúdo nacionais (incluindo roteiristas, diretores, atores, etc) e internacionais (distribuidores de filmes, séries e desenhos animados). Nesse sentido, tem-se um baixo poder de barganha no cenário brasileiro devido às poucas possibilidades de artistas televisionarem suas criações em emissoras de grande porte. Entretanto, quando se trata de empresas internacionais, o panorama muda, pois estas têm a possibilidade de vender para inúmeras emissoras,

plataformas de streaming e consumidores diretos por todo o globo. Portanto, nesse caso, seu poder de barganha aumenta consideravelmente.

Poder de barganha dos compradores: O cliente, para a Rede Gazeta, pode ser tanto a “audiência” quanto os “patrocinadores”. Do lado da audiência, percebe-se um aumento de poder de barganha nos últimos anos devido à polarização político-ideológica que o Brasil tem enfrentado. Assim, telespectadores cobram atitudes condizentes com suas opiniões por parte das emissoras e tendem a migrar para outros canais quando não recebem o que esperam. Já do lado dos patrocinadores, a TV deixou de ser a primeira opção de praça para muitos. Com os avanços da tecnologia e dos algoritmos de marketing segmentado, existem hoje muitos outros meios de atingir clientes com maior eficácia e menor capital investido. Assim, os anunciantes passam a ter um maior poder de barganha para com as emissoras.

Ameaça de serviços/produtos substitutos: A indústria televisiva tem encontrado agressiva competição nas plataformas de streaming e redes sociais. Cita-se, por exemplo, Netflix e Youtube como importantes concorrentes, atraindo 33% e 64% da população brasileira atualmente. Entretanto, dado que, segundo a pesquisa do Kantar IBOPE, a televisão ainda persiste em 97% dos lares nacionais e teve seu consumo intensificado durante a pandemia, infere-se que a ameaça de substitutos é mediana, pois os dados mostram que ela ainda domina o momento de lazer do brasileiro médio.

Rivalidade entre concorrentes: Historicamente, emissoras de TV aberta foram sempre muito competitivas, lutando por certos artistas e punindo aqueles que migravam para uma rival. Entretanto, esse cenário tem se atenuado. Hoje, já se vê artistas e atores passando da Globo para o SBT, ou do SBT para a Record, sem grandes polêmicas. Uma exceção é o caso Globo-Record, as quais ainda não compartilham artistas e apresentadores. Ainda assim, é fato que a competição por direitos de transmissão, especialmente esportivos, é acirrada, assim como por scores de audiência absolutos.

4.1.2. Análise SWOT

MATRIZ SWOT		
	Fatores Internos (Controláveis)	Fatores Externos (Incontroláveis)
Pontos Fortes	Forças <ul style="list-style-type: none"> - Rede televisiva com maior alcance no Espírito Santo; - 16 veículos de comunicação que se conectam todos os dias com os capixabas; - Afiliada da Rede Globo; - Alta qualidade de equipamentos técnicos e de filmagem. 	Oportunidades <ul style="list-style-type: none"> - Popularização de serviços on-demand para conteúdo televisivo, abrindo espaço para investimentos em streaming de programas da TV Gazeta; - Baixa incidência de conteúdo culturalmente capixaba para a população local.
Pontos Fracos	Fraquezas <ul style="list-style-type: none"> - Empresa tradicional; - Poucos funcionários focados na área de inovação; - Dependência de dados do People Meter, que não discriminam demográficos e grade de programação com acurácia satisfatória; - Limitados pelas decisões executivas da Rede Globo. 	Ameaças <ul style="list-style-type: none"> - Alta concorrência no mercado de comunicação, intensificada pelos artistas de redes sociais; - Preferência por outros aparelhos eletrônicos e mídias sociais para entretenimento por partes dos telespectadores; - Crescimento das plataformas de streaming; - Enfraquecimento da mídia tradicional (em especial noticiários) por conta da polarização informacional do país; - Certo desinteresse da população brasileira por noticiários incisivamente políticos.

4.1.3. Planejamento Geral da Solução

Para nosso projeto, recebemos um documento Excel com 18 abas, cada qual contendo taxas de audiência colhidas pelo PeopleMeter em diferentes períodos e divididas por gênero, faixa etária e classe socioeconômica.

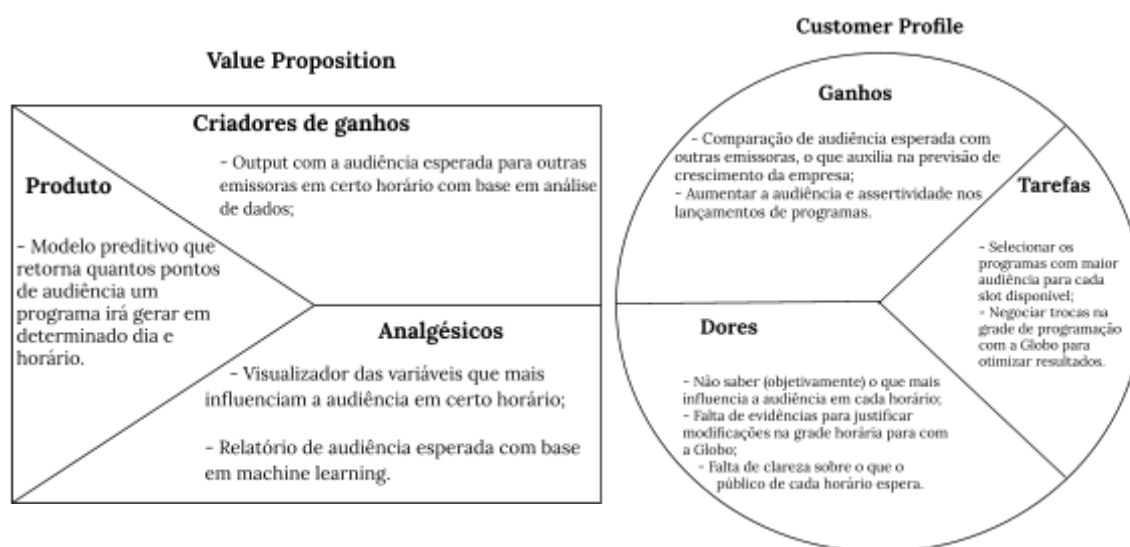
Com esse documento, pretendemos construir um modelo preditivo em que o usuário (funcionário da Rede Gazeta) possa inserir a data e horário desejados para um novo programa assim como o gênero do programa em questão. Com isso, o objetivo é prever o score de audiência nessa data e slot, indicando as variáveis que mais tiverem peso para o modelo para, assim, criar ações mais efetivas para maximizar o retorno financeiro do evento.

Nesse sentido, nosso produto trata de regressão, pois nós vamos retornar um valor dentre infinitas opções. O output secundário, no entanto, relacionado às variáveis utilizadas na predição, talvez seja classificatório. Ainda não temos como saber.

Além disso, em termos práticos, a solução proposta será utilizada pelas equipes de Programação e Marketing da Rede Gazeta. Desse modo, de acordo com os dados fornecidos por esse modelo, a Rede Gazeta pode definir um esforço (maior ou menor) necessário para divulgar certo programa, realizar trocas na grade de horário ou mesmo negociar slots com a Globo. Tudo isso, se bem aplicado, contribuirá para um maior retorno financeiro para os investimentos da empresa.

Por fim, o critério de sucesso ainda não foi bem definido, dado que não começamos a implementar o modelo. Entretanto, antecipamos que ele será determinado de acordo com uma tolerância máxima para o erro da predição. Isso, por sua vez, será aferido na fase de testes com um training set derivado dos dados que recebemos. Logo, consideramos o modelo preditivo um sucesso ele retorna uma margem de erro pequena ou desprezível.

4.1.4. Value Proposition Canvas



4.1.5. Matriz de Riscos

Matriz de Risco											
Probabilidade		Riscos					Oportunidade				
Muito Alta	5				Atrasos por conta de membros do grupo faltando						
Alta	4						Entregar análises fidedignas considerando as capacidades técnicas do grupo	Realizar os autoestudos antes das aulas e desenvolvimentos relacionados			
Médio	3			Parceiros de projeto com expectativa diferente de produto, superestimando	Concentração de tarefas em pessoas específicas do grupo	Nosso produto não se destacar em relação aos outros grupos	A solução ser adotada pela Rede Gazeta				Expandir a audiência da rede Gazeta
Baixa	2				Não conseguir entregar o projeto a tempo	Não conseguir analisar todos os dados e trazer vieses no modelo preditivo					
Muito Baixa	1			Grupo não se entrosar	Mudanças na LGPD, impactando as amostras que o aparelho pode coletar	Violação dos dados do Ibope. Não entregar previsões assertivas					
		1	2	3	4	5	5	4	3	2	1
		Muito Baixo	Baixo	Médio	Alta	Muito Alta	Muito Alta	Alta	Médio	Baixo	Muito Baixo
		Impacto									

† Matriz de risco - Grupo 5

4.1.6. Personas

Persona 1

Marta Lopes, 32 anos, Analista de Comunicação.

"Comunicativa, gosta de trabalhar com as pessoas e conhecer diferentes culturas"

Biografia: Migrou do Marketing para a área de Comunicação; Decidida, prática e ótima negociadora.

Dores/Motivações atuais com o problema: Não consegue negociar com os anunciantes de forma que correlacione o horário com a classe social que é alvo do anúncio.

Objetivos/necessidades específicas em relação ao problema: Focalizar os anunciantes para horários com maior potencial de compradores; Ter evidências para negociar taxas maiores com anunciantes fora do tradicional "horário nobre".

Persona 2

Giovanna Mattos, 28 anos, Gerente Geral de Marketing.

"Analítica, adora dados e quer transformar o ambiente em que trabalha através da tecnologia."

Biografia: Recentemente contratada e ainda se familiarizando com a empresa e suas particularidades; Experiência com Google Ads e análise de dados; Trabalhou como freelancer por bastante tempo.

Dores/Motivações atuais com o problema: Não há a possibilidade de prever a necessidade de reforço de campanhas de marketing direcionadas à horários; Falta de evidências para aprovar o orçamento de marketing.

Objetivos/necessidades específicas em relação ao problema: Alocar recursos de forma eficiente na divulgação de novos eventos, priorizando aqueles com menor potencial de audiência; Identificar fatores que mais contribuem para a audiência de um programa a fim de enfatizá-los na campanha de marketing; Ter evidências para negociar orçamentos maiores e promoções para o departamento de marketing, baseando-se comparação de audiência esperada e audiência consolidada após campanhas.

Persona 3

Rodrigo Souza, 35 anos, Gerente de Operação e Programação.

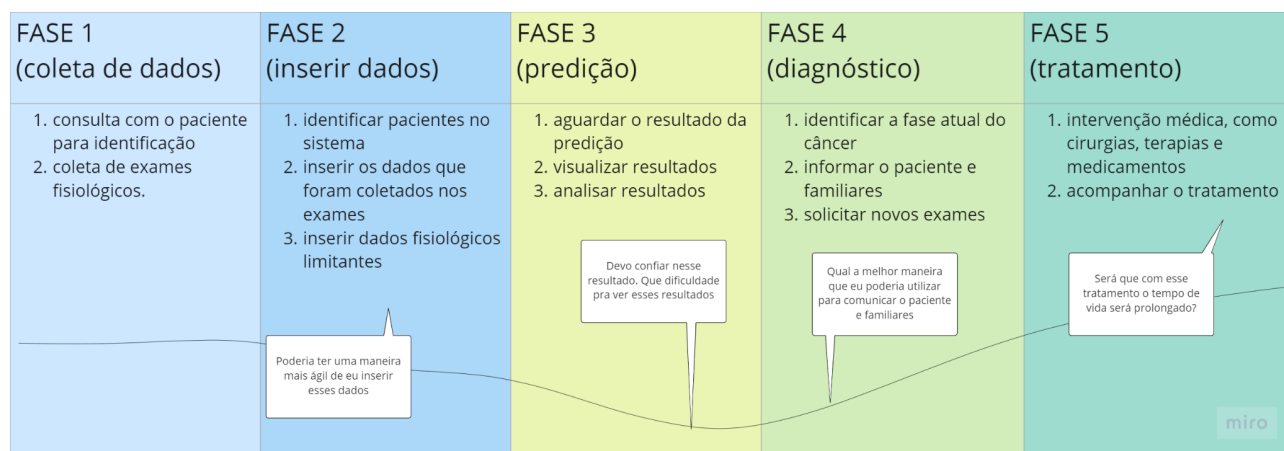
"Funcionário de longa-data desde o período de estágio. Enfrenta dificuldades para justificar os seus planos de ação de forma objetiva desde cedo."

Biografia: Funcionário de longa-data muito familiarizado com a grade atual; Opiniões fortes sobre os programas existentes e muitas ideias de melhoria; Intuitivo, tende a tomar decisões com base em suas impressões subjetivas.

Dores/Motivações atuais com o problema: Não saber (objetivamente) o que mais influencia a audiência em cada horário; Falta de evidências para justificar modificações na grade horária para com a Globo; Falta de clareza sobre o que o público de cada horário espera.

Objetivos/necessidades específicas em relação ao problema: Selecionar os programas com maior audiência para cada slot disponível; Identificar o que faz um programa ter sucesso em cada slot para produzir melhores eventos; Negociar troca de programação em certos slots com a Globo; Suprir expectativas do público de determinado slot.

4.1.7. Jornadas do Usuário



4.2. Compreensão dos Dados

Recebemos duas planilhas em formato XLSX que agrupam dados coletados pelo Kantar IBOPE Media a partir de um aparelho de depuração de audiência chamado “People Meter”. Esse aparelho torna possível registrar o perfil de quem está assistindo com os demográficos de que faz parte, o número total de indivíduos atingidos e o tempo de consumo.

O primeiro arquivo, denominado “TV_Histórico”, contém 18 abas elencando dados de audiência por emissora e dia da semana. Mais especificamente, cada emissora é dividida em três abas: uma para dias úteis (156.673 linhas), uma para sábado (31.105 linhas) e uma para domingo (também 31.105 linhas). O período analisado, por sua vez, é de 6 de junho de 2020 a 25 de junho de 2022, sendo que cada dia é tabelado de cinco em cinco minutos. Uma peculiaridade, nesse sentido, é que o horário a cada dia começa em 24:00:00 (correspondendo à meia-noite) e vai até 29:55:00 (correspondendo às 5:55:00 da manhã); depois, a contagem se reinicia às 6:00:00.

Cada aba contém as porcentagem de três métricas de audiência: Rating (considerado o score de audiência), Fidelidade (parcela dos televisores que ficou no mesmo canal por pelo menos um minuto), Reach (parcela de televisores alcançada) e Share (parcelas de televisores sintonizados com certa emissora dentre todos os ligados em dado momento). Esses valores são, por sua vez, subdivididos em classe socioeconômica (AB, C1, C2 e DE), sexo (masculino e feminino) e idade (4-11 anos, 12-17 anos, 18-24 anos, 25-34 anos, 35-49 anos, 50-59 anos e 60+ anos).

Já o segundo arquivo, “grade_Diária_06_2020_a_06_2022.xlsx”, traz uma coluna “Praça”, correspondente ao local de transmissão, “Data”, “Faixa Horária” e uma coluna para cada emissora, contendo a programação para cada slot de junho de 2020 a junho de 2022.

Para fins de análise de dados, esse documento foi agregado com o primeiro através da adição de duas colunas (“Programa” e “Gênero”) seguindo a paridade de horário entre as tabelas. Além disso, os horários foram formatados em intervalos de 30 minutos através de média aritmética para repetição. O campo de data também foi dividido em colunas de ano, mês e dia.

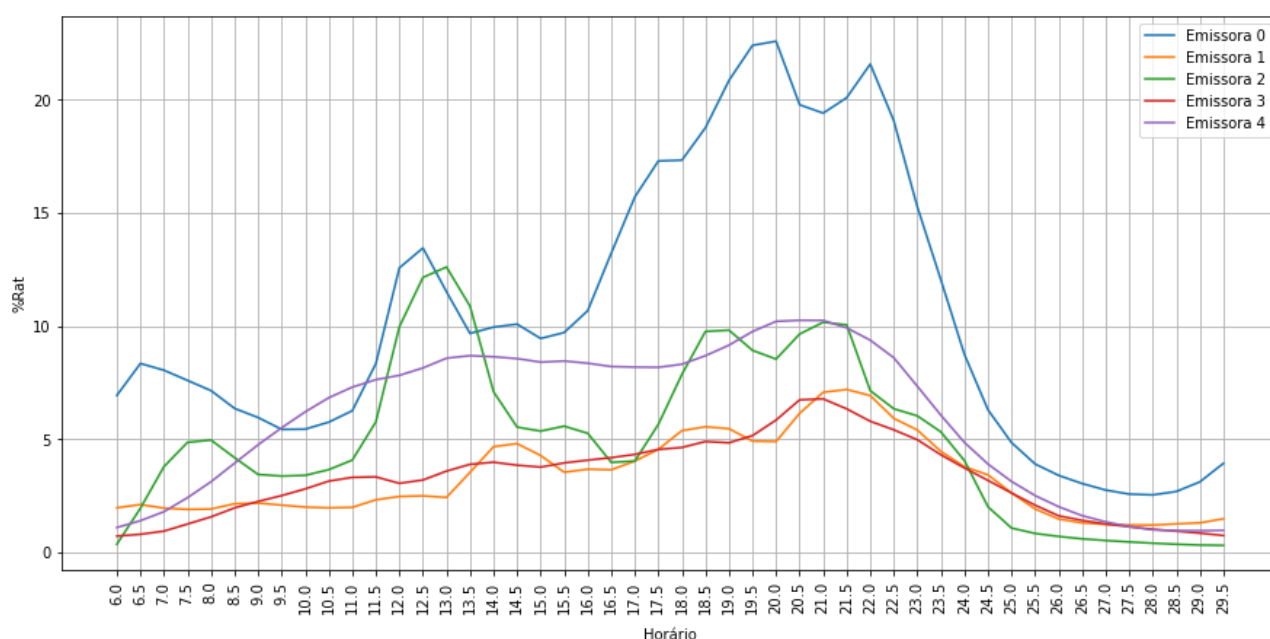
Ademais, é importante notar que não há contingências em relação à qualidade e preservação dos dados, ou seja, células vazias ou “NaN” não são um problema até o momento.

Por fim, os dados disponibilizados pelo Kantar IBOPE são de natureza sigilosa, de modo que o nome das emissoras não deve ser mencionado em qualquer documento público.

Análise preliminar dos dados

Para todas as análises, foram consideradas apenas as audiências em Rat, conforme recomendação do cliente. Isso se dá porque o Rat é justamente um cálculo mais significativo sobre os valores de Fid e Shr.

Audiência por hora por emissora (Seg a Sex)



Este gráfico apresenta cada emissora como uma linha e mostra a audiência de Rat de domicílios totais a cada hora do dia durante a semana. Nesse sentido, há uma dominância clara da Emissora 0 durante todo o período, exceto das 9h30 às 11h15, quando a Emissora 4, representando itens não identificados, como streaming, a supera. Uma hipótese que explica esse comportamento é o fato de este ser o horário preferido por donos(as) de casa para realizar tarefas domésticas; nesse contexto, é provável que elas utilizem os desenhos animados disponíveis em plataformas de streaming para distrair as crianças, segundo insight do parceiro de projetos.

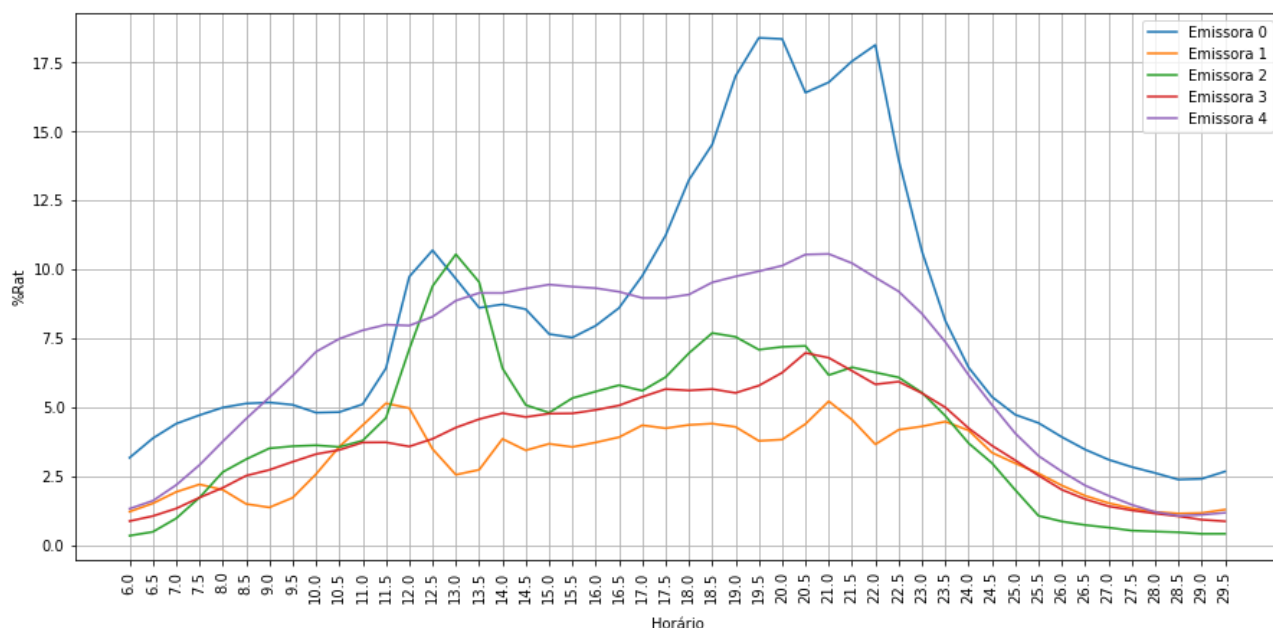
Ademais, o primeiro pico relativo ocorre às 6h30. Sugere-se que isso ocorra devido à veiculação do primeiro evento jornalístico do dia justamente no momento em que a maioria das pessoas acorda para ir à escola ou ao trabalho. Isso volta a ocorrer ao meio-dia, aproveitando a atenção das pessoas em seu horário de almoço. Aqui, há uma competição acirrada entre a Emissora 0 e a Emissora 2, dado que ambas apresentam o segundo jornal do dia nesse horário. Percebe-se, portanto, que a audiência se divide entre as duas, o que indica a necessidade de se diferenciar e segmentar o conteúdo a

fim de melhor atender a população capixaba. Uma possibilidade seria dedicar mais tempo a assuntos locais e temas culturais para gerar mais identificação no público, instigando um sentimento de orgulho e pertencimento que os fará voltar para o programa dia a dia.

A partir daí, há uma baixa de audiência durante o final do bloco jornalístico e bloco de filmes, provavelmente por ser um horário em que a maioria das pessoas está ocupada com suas tarefas diárias, profissionais ou não.

A audiência volta a subir no final da tarde com a transmissão das primeiras telenovelas e aumenta consistentemente até cerca das 20h, quando começa o principal programa jornalístico do dia. Nesse momento, há uma queda súbita com grandes chances de ser motivada por divergências ideológicas. Por fim, o último pico se dá às 22h30, em que ora passam filmes, ora reportagens, ora reality shows populares.

Audiência por hora por emissora (Sábado)



Aos sábados, nota-se uma configuração relativa semelhante de audiência entre as emissoras. Porém, ao tratar de valores absolutos, percebe-se que as porcentagem de Rat são claramente menores aos sábados, chegando à máxima geral de 18 contra a de 21 durante a semana.

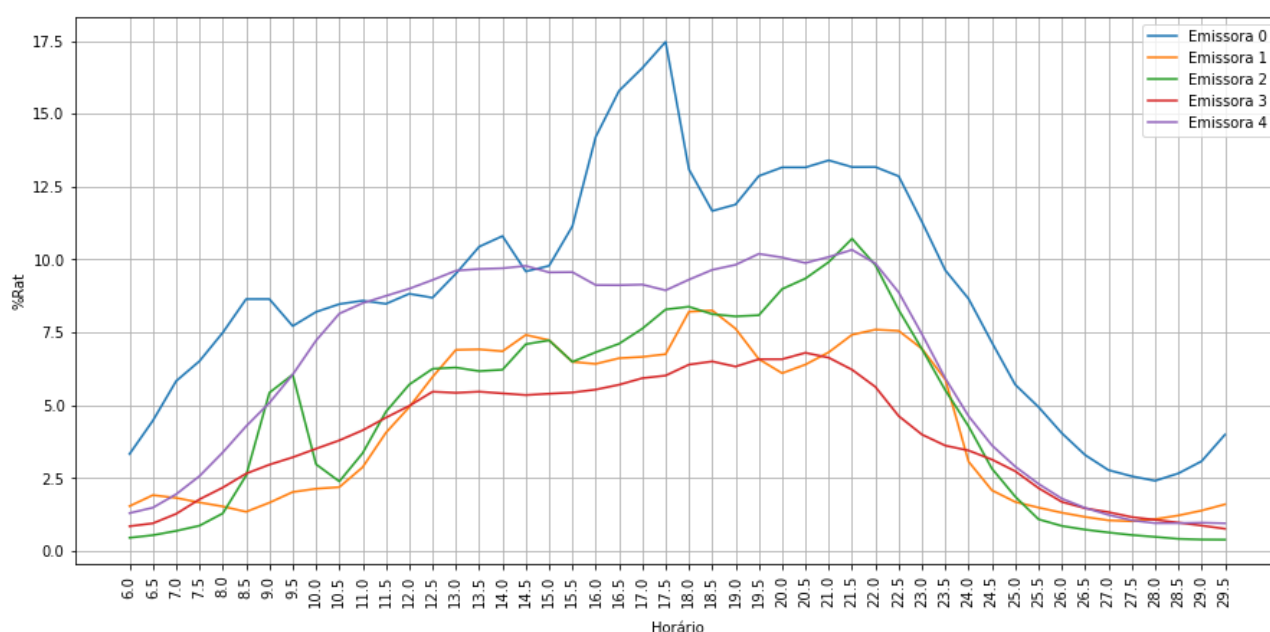
Ademais, dentre outras particularidades, há uma presença considerável da Emissora 4, representando plataformas de streaming e outros conteúdos não identificados pelo PeopleMeter. Isso é visível pela primeira vez das 8h45 até as 11h30. Nossa hipótese, nesse caso, segue a linha da anterior em que pais e guardiões utilizam essas ferramentas para distrair as crianças durante o período da manhã. Outra explicação é um maior consumo de entretenimento durante o período da manhã em finais de semana, em que a família se reúne para tomar café e

pode compartilhar conteúdos no tablet e no celular. Já os picos na hora do almoço seguem a mesma lógica do resto da semana, ainda que com valor absoluto inferior a 20%.

Durante a tarde, a Emissora 4 novamente sobe em audiência comparativamente, apesar de se manter no mesmo valor absoluto dos dias úteis. Isso indica que as pessoas assistem a conteúdo de streaming nesse horário mais consistentemente do que à TV aberta, isto é, em seu tempo de descanso, elas preferem streaming aos filmes e shows de auditório característicos dos sábados.

Por fim, o resto do dia segue o comportamento de segunda a sexta.

Audiência por hora por emissora (Domingo)



Domingo é um dia atípico, como pode ser claramente visto pelas flutuações nas linhas. O dia já começa com audiência considerável e consistente das 8h30 às 9h para a Emissora 0, que veicula eventos rurais nesse período. Um fato curioso é que a queda da Emissora 0 entre 9h e 9h30 se alinha quase que perfeitamente com a subida da Emissora 2, indicando uma migração dos telespectadores para assistir à programação mais decididamente capixaba dessa emissora para esse slot.

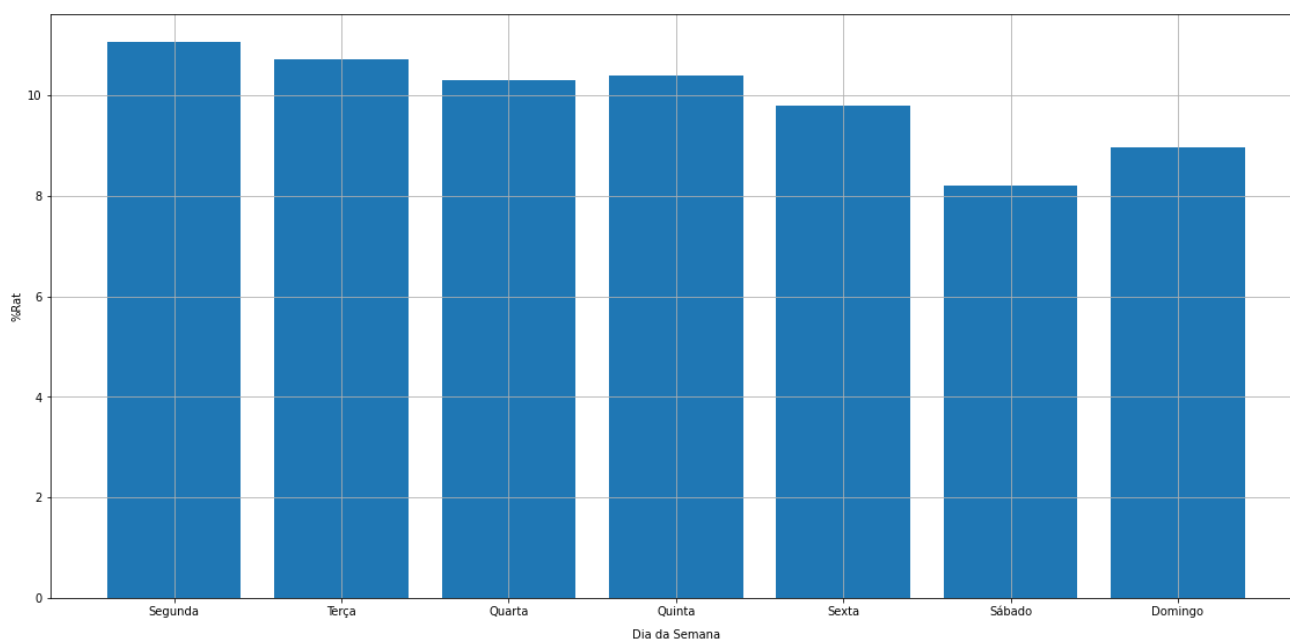
No horário do almoço, os picos não ocorrem mais devido à ausência de eventos jornalísticos. Ainda assim, o leve aumento às 13h30 para a Emissora 0 mostra que a população gosta de sentar para assistir a filme – conteúdo comum para esse período aos domingos – após se alimentar.

O pico principal acontece mais cedo, às 17h30, quando começam os programas de auditório mais populares de cada emissora. Além disso, existe um público cativo consistente para a programação das 20h às 23h, quando são transmitidos blocos de reportagens e reality

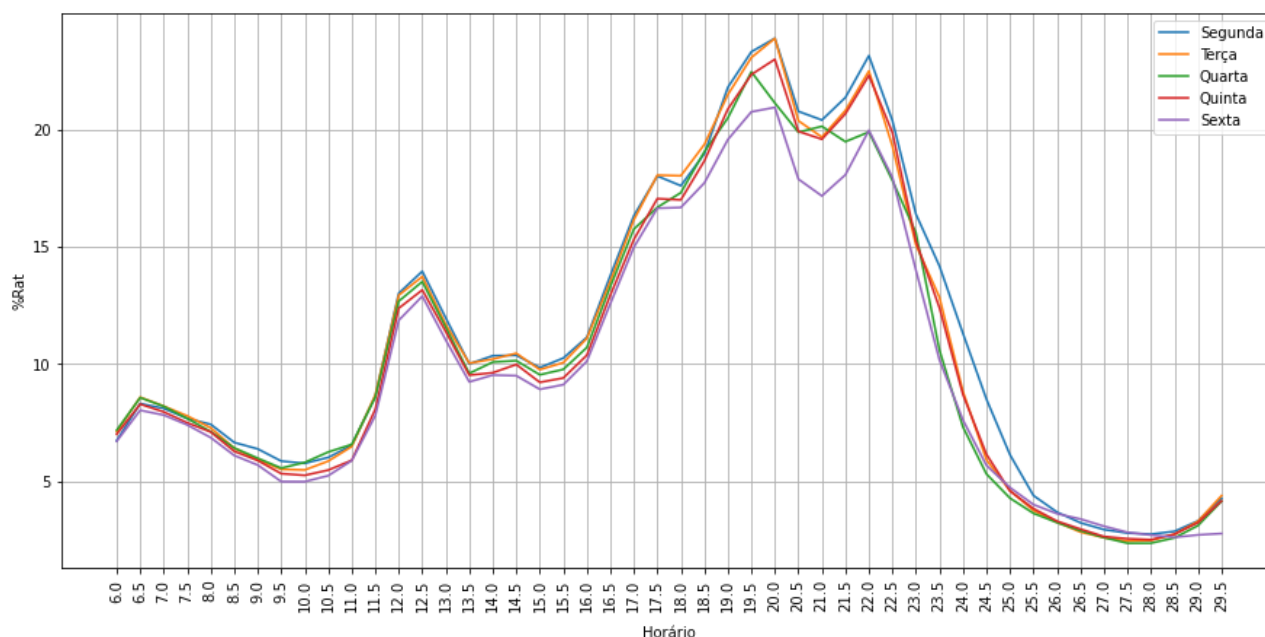
shows. Ainda assim, em segundo nível, a Emissora 4 ainda alcança uma audiência competitiva com seus eventos de auditório mais tradicionais.

Audiência da Emissora 0 por dia da semana

Dada a dominância geral da Emissora vista nos gráficos anteriores, decidimos seguir as análises apenas com a audiência dela, pois outras emissoras não apresentam valores significativos em comparação.



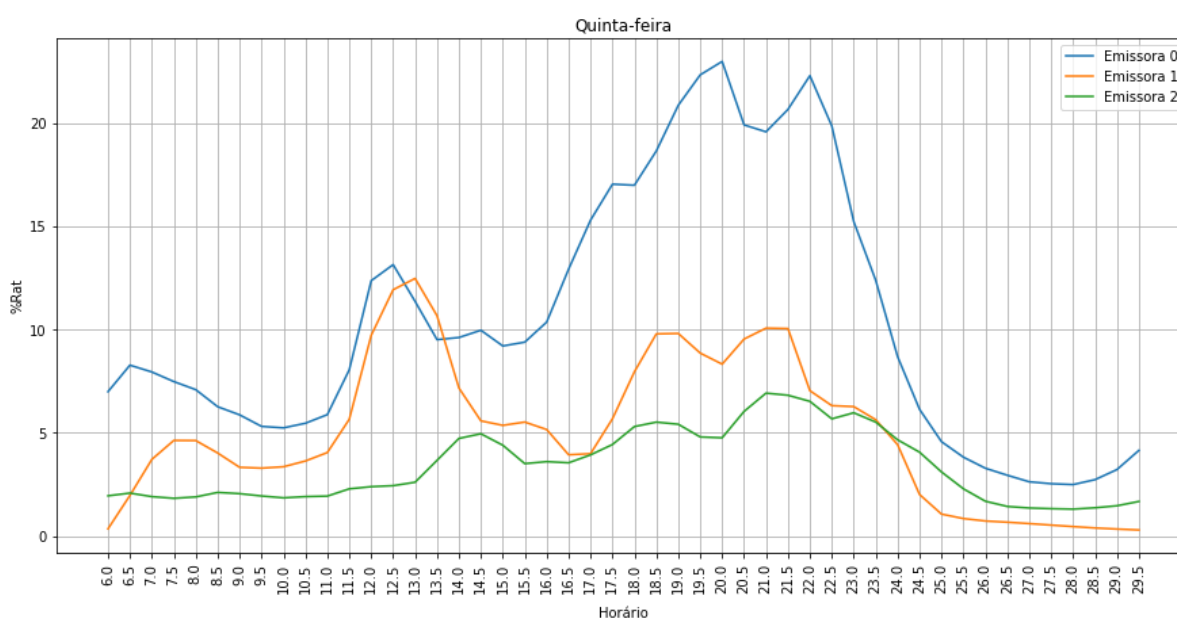
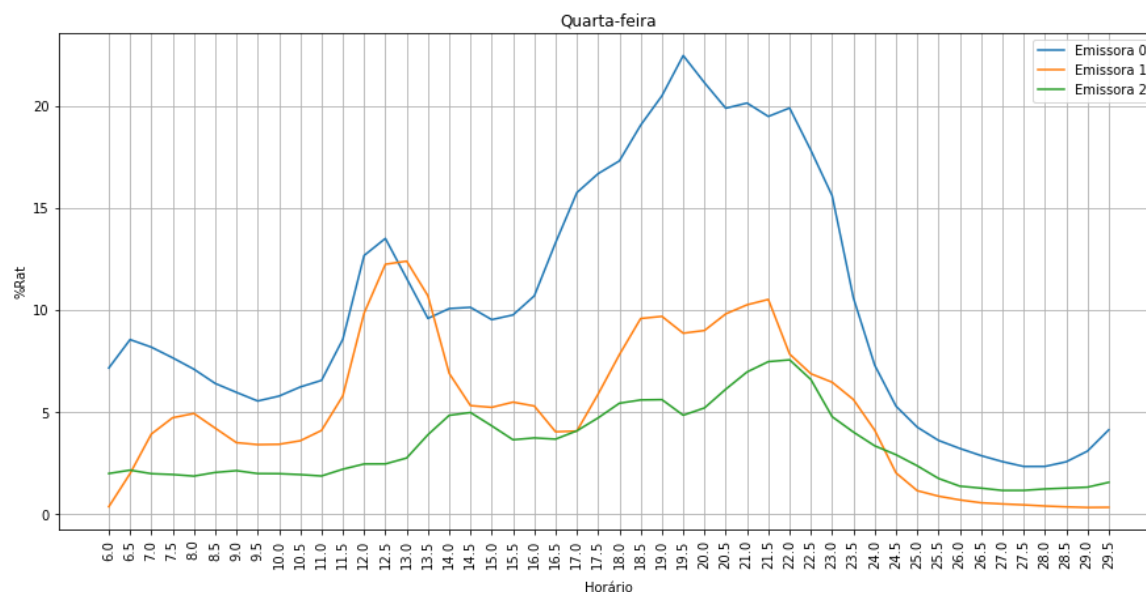
A primeira coisa a se notar é que as audiências para cada dia, exceto sábado e domingo, apresentam pouca variabilidade, ficando sempre em torno de 10-10.5 Rat%. Nesse sentido, as baixas no final de semana são fáceis de explicar: as pessoas têm mais tempo livre nesses dias e buscam outras formas de lazer além da televisão. Por outro lado, explicar por que mais pessoas sintoniza às segundas tem se provado um desafio. A programação semanal em certo mês tende a se manter regular, de modo que não grandes mudanças entre segunda e terça, por exemplo. Para aumentar nosso entendimento nesse aspecto, geramos também os gráficos de horário para cada dia da semana, a fim de visualizar quais blocos e, consequentemente, quais programas mais atraíam os telespectadores.



Como sugerido, as audiências são muito semelhantes durante a semana. Ainda assim, olhando com mais atenção, percebe-se que os valores para sexta-feira tendem a cair cada vez mais conforme o dia progride, decerto porque as pessoas preferem sair após o trabalho nessas noites em antecipação do final de semana.

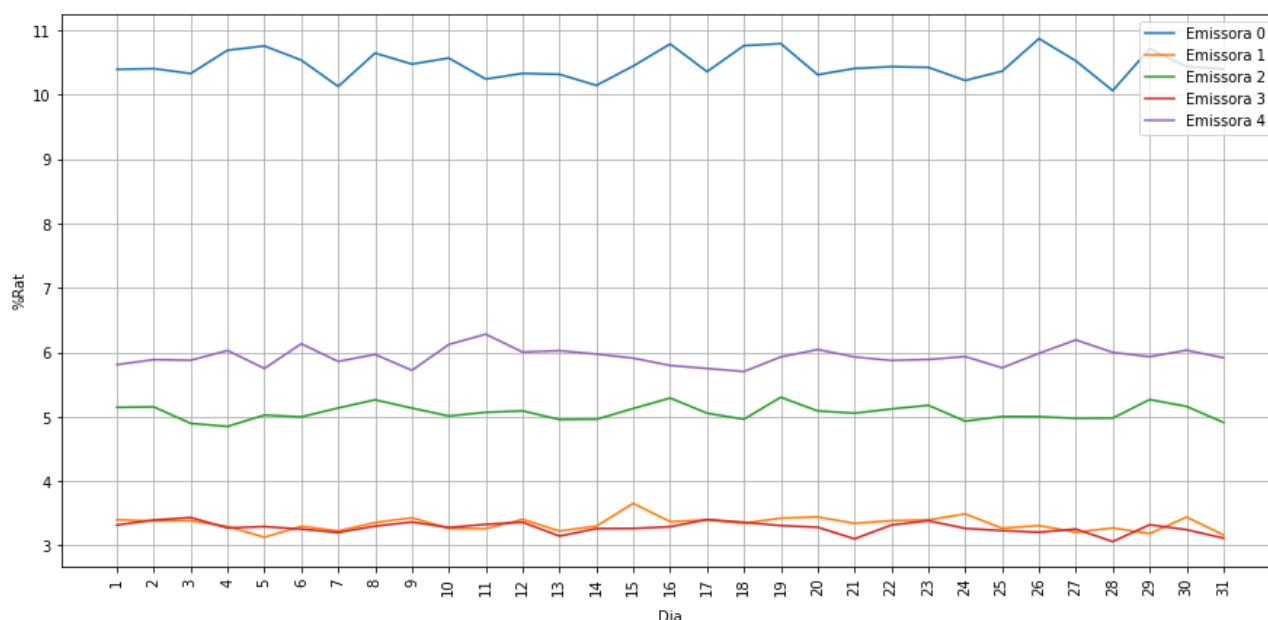
Já a terça-feira segue o padrão de segunda com muita precisão, exceto no final da noite e início da madrugada. Um ponto curioso aqui é que, pelo jeito como as horas são apresentadas no gráfico, temos a impressão de que a hora 29.0 no gráfico de segunda-feira representa a madrugada de terça-feira. Porém, na realidade, esse horário se refere à madrugada entre domingo e segunda-feira, pois, na planilha, todos os dias iniciam com horário 24.0 e só depois retornam para 6. Com esse insight, percebemos que um motivo para a maior audiência na madrugada de segunda-feira é simplesmente o fato de esta ser uma “extensão” de domingo, de modo que a muitas pessoas ainda não voltaram a seu horário de sono habitual e ficam acordadas até mais tarde.

Ademais, para examinar as baixas relativas nas quartas e quintas, plotamos também as audiências por horário para cada um desses dias considerando também outras emissoras. O objetivo era identificar se havia migração por algum programa específico.



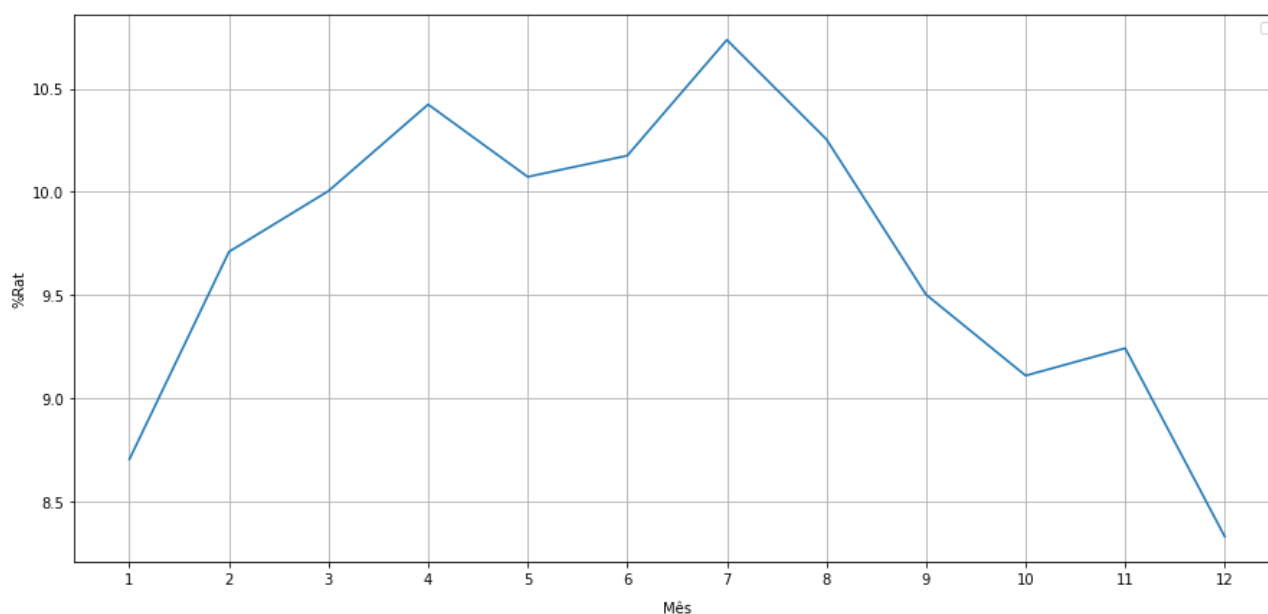
Ambos os dias, entretanto, apresentam flutuações semelhantes às dos outros dias da semana, como aferido pela média semana no primeiro gráfico mostrado nessa análise. Assim, ainda nos falta informação para induzir especificamente o que pode influenciar esse comportamento diferenciado às quartas-feiras e, em menor grau, às quintas.

Audiência por dia do mês por emissora



Novamente, a dominância da Emissora 0 é clara, e as audiências flutuam em um intervalo muito pequeno. Nesse sentido, ao comparar as datas dos picos relativos com a programação vigente, há resultados inconclusivos. Em alguns casos, realmente houve eventos especiais, como jogos de futebol, reality shows e blocos de reportagens populares. Em outros, no entanto, não há explicação clara do porquê esse dia em particular atrairia mais público. Por isso, deduzimos que, salvo raras exceções, não há muita variação de comportamento com o passar do mês.

Audiência da Emissora 0 por mês

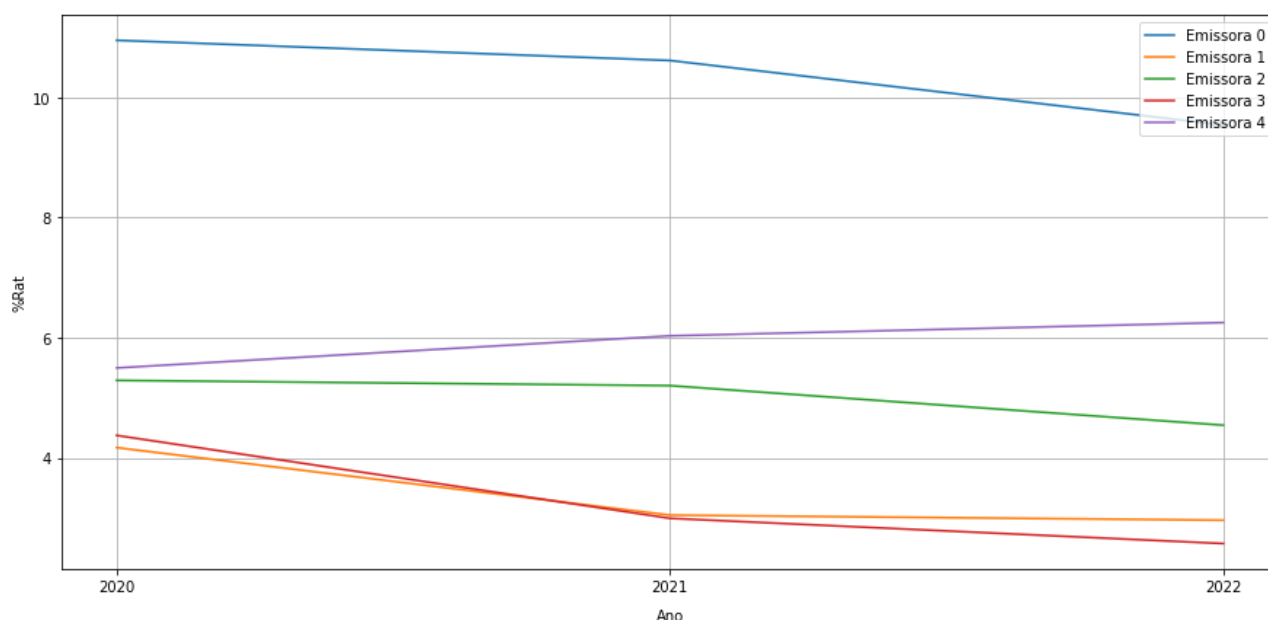


Há dois picos importantes na análise mensal de audiência. O primeiro ocorre em abril, provavelmente porque esse foi o mês em que o reality show mais popular da Emissora 0 foi televisionado nos últimos dois anos. Já o segundo se dá em julho, decerto pela veiculação das últimas Olimpíadas nesse período.

A queda que se segue pode ser explicada, em parte, pela usual troca de novelas e séries no segunda semestre. Nesse contexto, vê-se que a audiência ainda se mantém relativamente alta no mês de agosto, talvez pelo interesse da população nas novidades, mas diminui com o passar dos meses quando essa novidade inicial se esvai.

Há um leve aumento em novembro, que nós atribuímos como hipótese à finalização de atividades profissionais e escolares. É o momento em que as responsabilidades diminuem e as pessoas começam a tirar férias. Já em dezembro, com os planejamentos para as festas e possíveis viagens e outras formas de lazer presencial, a audiência volta a cair.

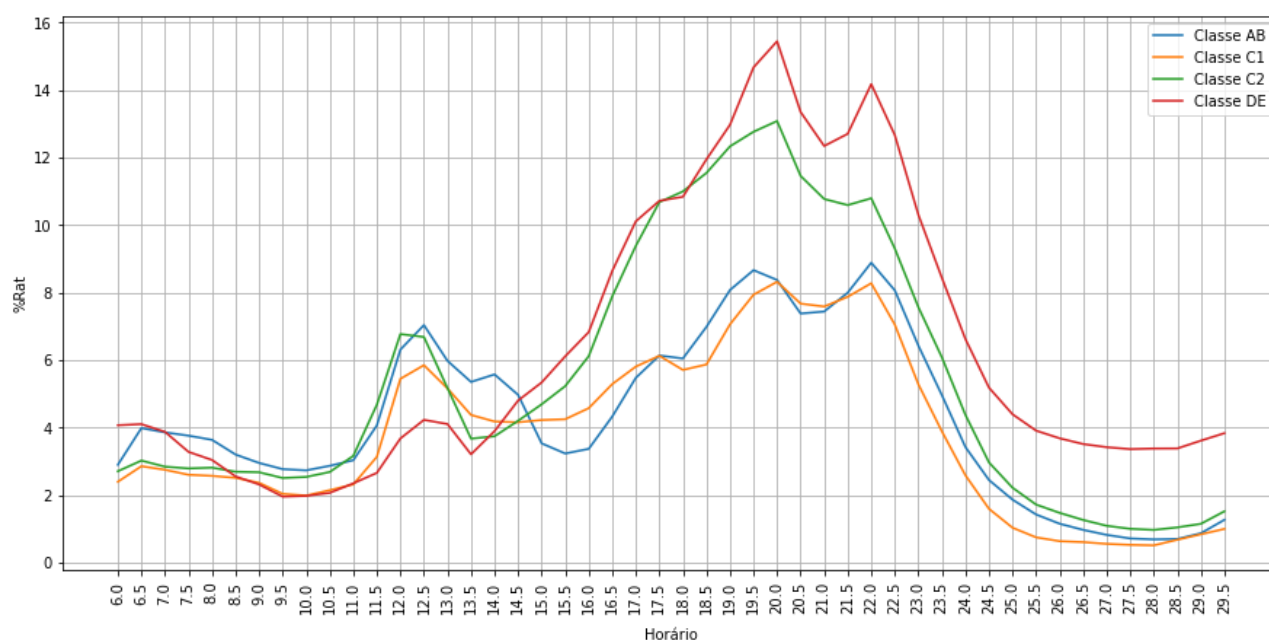
Audiência por emissora por ano



Nota-se uma leve queda de audiência da Emissora 0 de 2020 para 2022. Isso é explicado pela normalização pós-pandemia. Durante a crise de COVID-19, as pessoas ficaram muito mais tempo em casa e, portanto, tiveram mais oportunidades de assistir à televisão, além de terem mais interesse nas notícias quanto ao status da doença. Com a reabertura comercial, os níveis de audiência voltaram para o que acreditamos serem níveis pré-pandemia. Entretanto, como não temos dados de anos anteriores, não podemos validar essa hipótese.

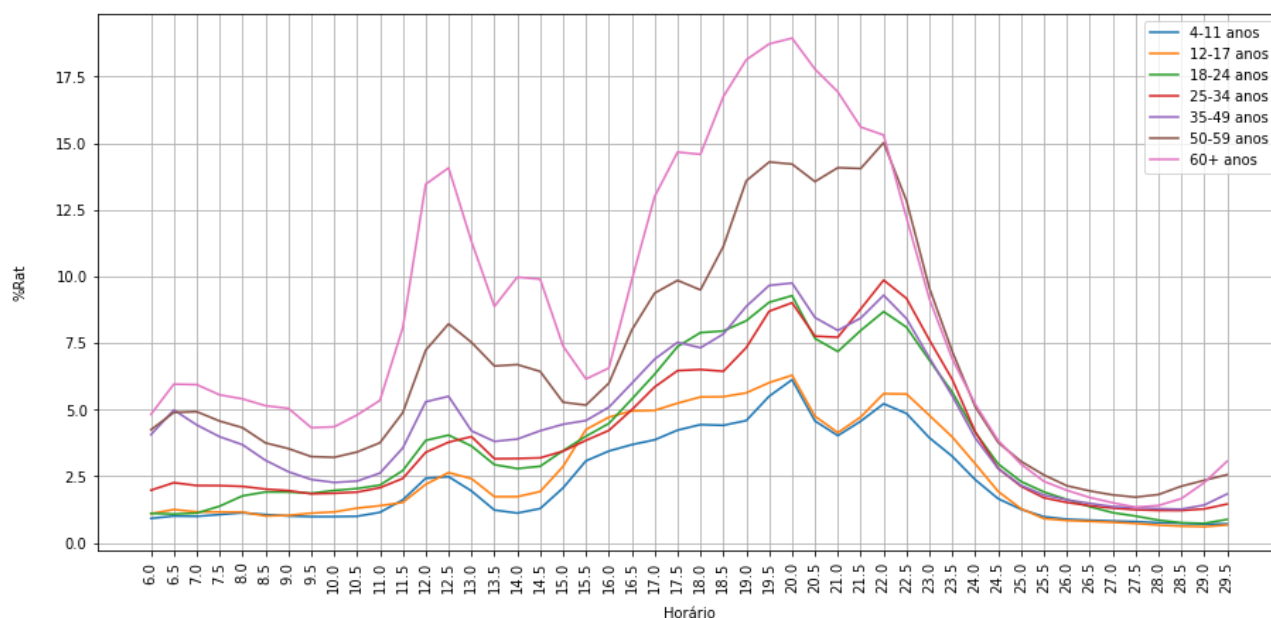
Percebe-se também que a Emissora 4, representando plataformas de streaming, teve certo crescimento nos últimos anos. Isso é corroborado por pesquisas que mostram que o consumo de streaming aumentou em 34% com a criação de mais conteúdo exclusivo ("State of Mobile 2022 - data.ai", 2022).

Audiência da Emissora 0 por classe socioeconômica



Na análise socioeconômica, o pico geral pertence à classe DE no bloco jornalístico do período da noite. Assim, deduz-se que essa classe tem preferência pelo jornalismo tradicional, enquanto as outras classes consomem uma programação mais variada e menos política, com programas de auditório, jornais locais e novelas, durante a manhã e tarde.

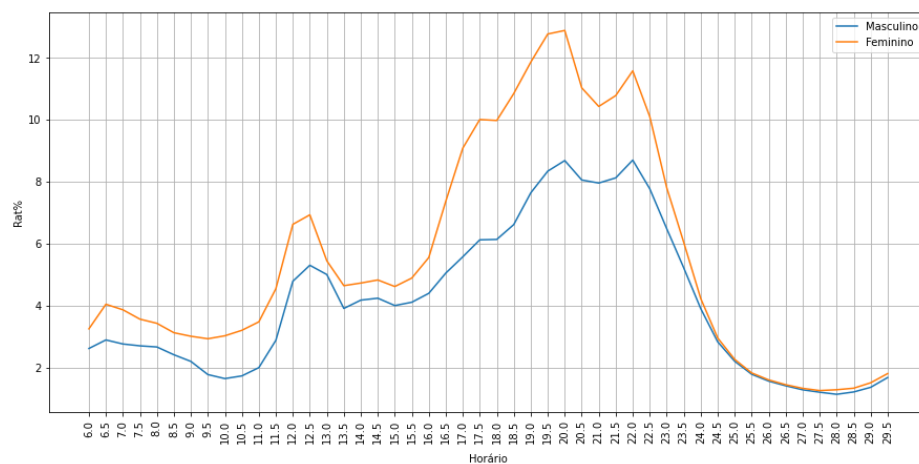
Audiência da Emissora 0 por faixa etária



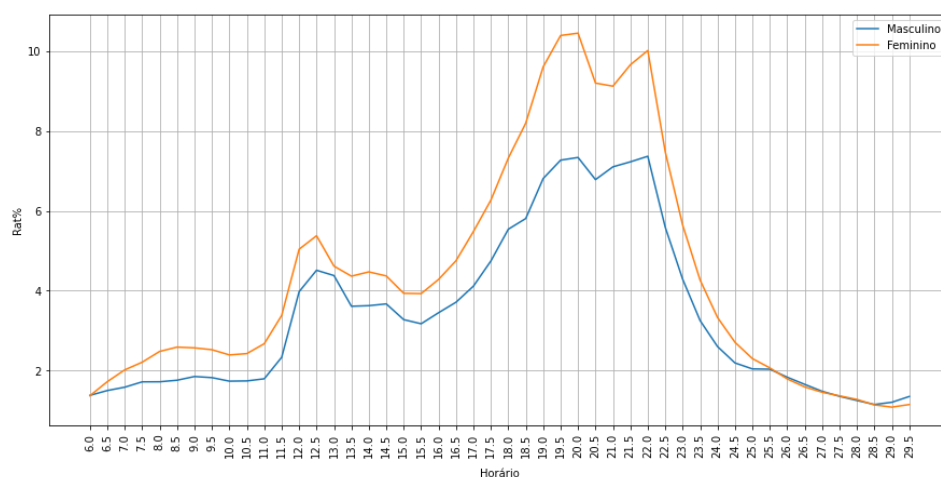
No que tange à faixa etária, a terceira idade domina em quase todos os casos e especialmente durante a manhã, provavelmente pela incidência de aposentadoria e horários mais flexíveis, além da tendência a acordar mais cedo (AMANDA CHAN 12 APRIL 2011, [s.d.]). Apesar disso, todas faixas apresentam comportamento e flutuações parecidas, ainda que em níveis diferentes, indicando que a categoria dos eventos a cada faixa horária agrada a população em geral de forma consistente. Um ponto interessante é que crianças aparecem com muito menos frequência, provavelmente pela escola e preferência por Youtube e outras plataformas.

Audiência total da Emissora 0 por gênero

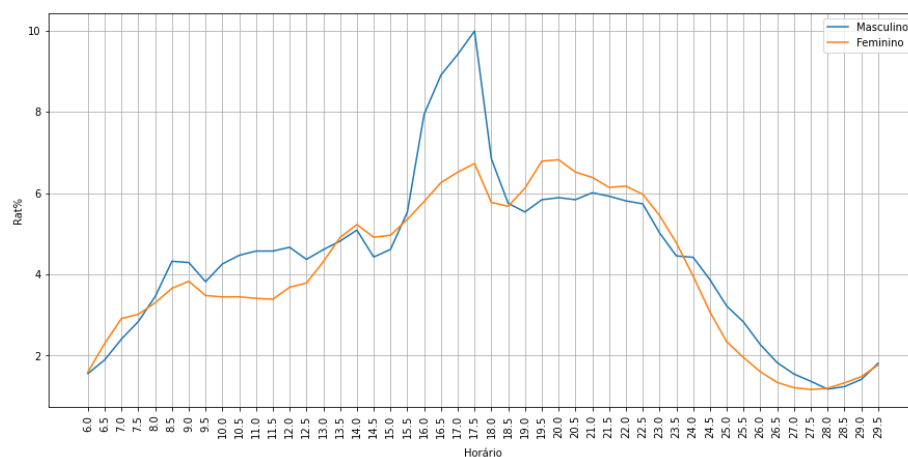
Segunda a sexta



Sábado

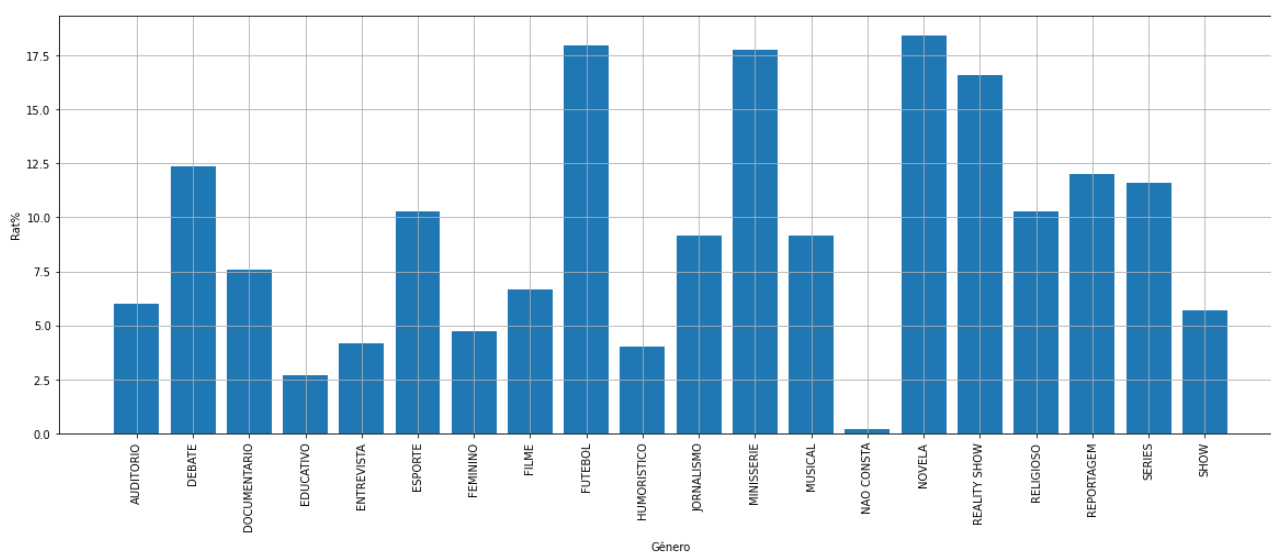


Domingo



A análise desses três gráficos mostra que, em geral, mulheres predominam na audiência geral por cerca de 1.5 Rat% durante todo o dia e quase 4 pontos durante o horário de pico. Esse resultado é surpreendente, dado que hoje apenas 7% das mulheres se declaram donas de casa e têm acesso a uma televisão o dia inteiro (“Parcela da população que se declara dona de casa cai para 7% em 26 anos”, 2019). Outra hipótese não validada é um maior interesse feminino por telenovelas. A exceção é o dia de domingo, quando homens ultrapassam as mulheres no horário de jogos de futebol.

Audiência da Emissora 0 por gênero



Geramos esse gráfico com o objetivo de entender quais gêneros eram mais assistidos. Nesse sentido, definimos uma lista de “preferidos” pelo público com base nos picos relativos, chegando aos itens de Debate, Futebol, Minissérie, Novela, Reality Show, Reportagem e Séries. Curiosamente, o gênero Jornalismo, do qual esperávamos uma das maiores audiências, não figura nessas máximas relativas. Isso é tão inusitado para nós que suspeitamos que talvez o PeopleMeter não esteja classificando todos os blocos jornalísticos corretamente, dado que são eles que atraem as maiores taxas de audiência nas análises por hora.

Considerações sobre o resultado desejado

O resultado desejado da predição (saída) tem dois componentes. O primeiro e principal é o score de audiência esperado para um evento de determinadas características em certo dia e horário. Por ser um número dentre inúmeras opções, sua natureza é contínua (float). O segundo output planejado é uma lista das variáveis consideradas para chegar ao resultado principal e o peso de cada uma delas. Até o momento, não ficou claro se esse resultado será classificatório, partindo do princípio de que nem todas as variáveis possíveis serão utilizadas em cada predição, ou contínuo, caso os pesos de cada variável possam mudar de predição em

predição. Será necessário estudar mais sobre modelos preditivos e fazer as primeiras implementações para sanar essa dúvida.

4.3. Preparação dos Dados

Otimização dos arquivos

O documento original, “TV_Histórico.xlsx”, continha cerca de um milhão de linhas quando somadas todas as suas abas. Logo, o carregamento desse arquivo no código demandava muito tempo e recursos computacionais, além de gerar dataframes mais difíceis de manipular. Portanto, nossa primeira ação em preparar os dados foi separar esse documento Excel em dezoito arquivos CSV, um para cada aba. Fizemos isso para ter um acesso mais assertivo aos dados desejados e, também, diminuir o overhead do algoritmo através de um formato mais leve que o XLSX.

Agregando emissoras

Após isso, para agregar informações semelhantes e facilitar a análise por emissora, juntamos os CSVs de dias úteis, sábado e domingo em um único CSV por canal. Em mais detalhes, dado que com a etapa um conseguimos três arquivos por emissora (“Seg a Sex”, “Sáb” e “Dom”), decidimos, nesta fase, juntar esses arquivos em um só, ordenados cronologicamente, para cada emissora.

Agregando emissoras

Após isso, para agregar informações semelhantes e facilitar a análise por emissora, juntamos os CSVs de dias úteis, sábado e domingo em um único CSV por canal. Em mais detalhes, dado que com a etapa um conseguimos três arquivos por emissora (“Seg a Sex”, “Sáb” e “Dom”), decidimos, nesta fase, juntar esses arquivos em um só, ordenados cronologicamente, para cada emissora.

Checando a existência de outliers

Checando valores ausentes

Descreva as etapas realizadas para definir os dados e os atributos descritivos dos dados (“features”) a serem utilizados. Essa descrição deve ser feita de modo a garantir uma futura reprodução do processo por outras pessoas, e deve conter:

- Descrição de quaisquer manipulações necessárias nos registros e suas respectivas features.
- Se aplicável, como deve ser feita a agregação de registros e/ou derivação de novos atributos.
- Se aplicável, como devem ser removidos ou substituídos valores ausentes/em branco.
- Identificação das features selecionadas, com descrição dos motivos de seleção.

Não deixe de usar tabelas e gráficos de visualização de dados para melhor ilustrar suas descrições.

IMPORTANTE: Crie tópicos utilizando a formatação “Heading 3” (ou menor) para que o Google Docs identifique e atualize o Sumário (é necessário apertar o botão Refresh no Sumário para ele coletar as atualizações)

4.4. Modelagem

Para a Sprint 3, você deve descrever aqui os experimentos realizados com os modelos (treinamentos e testes) até o momento. Não deixe de usar equações, tabelas e gráficos de visualização de dados para melhor ilustrar seus experimentos e resultados.

Para a Sprint 4, você deve realizar a descrição final dos experimentos realizados (treinamentos e testes), comparando modelos. Não deixe de usar equações, tabelas e gráficos de visualização de dados para melhor ilustrar seus experimentos e resultados.

4.5. Avaliação

Nesta seção, descreva a solução final de modelo preditivo, e justifique a escolha. Alinhe sua justificativa com a seção 4.1, resgatando o entendimento do negócio e explicando de que formas seu modelo atende os requisitos. Não deixe de usar equações, tabelas e gráficos de visualização de dados para melhor ilustrar seus argumentos.

4.6 Comparação de Modelos

5. Conclusões e Recomendações

Escreva, de forma resumida, sobre os principais resultados do seu projeto e faça recomendações formais ao seu parceiro de negócios em relação ao uso desse modelo. Você pode aproveitar este espaço para comentar sobre possíveis materiais extras, como um manual de usuário mais detalhado na seção “Anexos”.

Não se esqueça também das pessoas que serão potencialmente afetadas pelas decisões do modelo preditivo, e elabore recomendações que ajudem seu parceiro a tratá-las de maneira estratégica e ética.

6. Referências

State of Mobile 2022 - data.ai. Disponível em

< <https://www.data.ai/en/go/state-of-mobile-2022/> > .

AMANDA CHAN 12 APRIL 2011. Discovery Reveals Why Old People Go to Bed Early. Disponível em:

< <https://www.livescience.com/13666-older-people-sleep-wake-early.html> > .

Parcela da população que se declara dona de casa cai para 7% em 26 anos. Disponível em:

< <https://www1.folha.uol.com.br/mercado/2019/08/parcela-da-populacao-que-se-declara-dona-de-casa-cai-para-7-em-26-anos.shtml> > . Acesso em: 14 ago. 2022.

Anexos

Utilize esta seção para anexar materiais como manuais de usuário, documentos complementares que ficaram grandes e não couberam no corpo do texto etc.