

Apresentação para vaga de Data Analyst - In Loco

por Victor Félix Pimenta





Introdução

- Material
 - Base de dados de emplacamento de automóveis
 - Informações sobre marcas, modelos, número de emplacamentos por data e tipos de automóvel
- Objetivo
 - Identificar problemas e oportunidades nos dados
- Ferramentas utilizadas
 - Script desenvolvido na linguagem R
 - Uso das bibliotecas: `base`, `readxl`, `tseries` e `optparse`
 - Implementação disponível em <https://github.com/vfpimenta/auto-project>



Informações

- 18 marcas de automóveis
 - GM, HYUNDAI, FORD, RENAULT, TOYOTA, JEEP, FIAT, HONDA, VOLKSWAGEN, NISSAN, PEUGEOT, CITROEN, KIA, MERCEDES BENZ, AUDI, MITSUBISHI, BMW e LAND ROVER
- 80 modelos de automóveis
- 12 tipos de automóveis
 - HATCH PEQUENO, VEICULO DE ENTRADA, SEDAN PEQUENO, SEDAN MEDIO, SUV, MONOCAB, GRANDCAB, SEDAN COMPACTO, SW MEDIO, HATCH MEDIO, SEDAN GRANDE e DESCONHECIDO (adaptado de NA)



Problemas

- Algumas marcas estão com problemas de digitação
 - Esses valores foram corrigidos durante a leitura dos dados
- Alguns modelos não têm tipo demarcado
 - Esses modelos foram marcados como DESCONHECIDO e retirados das análises de forma a não comprometer os demais dados
- Existem vários valores faltando quanto ao número de emplacamento
 - Optou-se por não remover ou preencher esses dados artificialmente devido ao problema seguinte, em vez disso foram tomadas precauções no processo de análise
- Há poucos dados disponíveis
 - Por se tratarem de séries bem curtas e não se ter mais detalhes sobre onde e como foram coletados os dados, preferiu-se não gerar dados artificiais e promover uma leitura mais criteriosa, assim como no ponto anterior



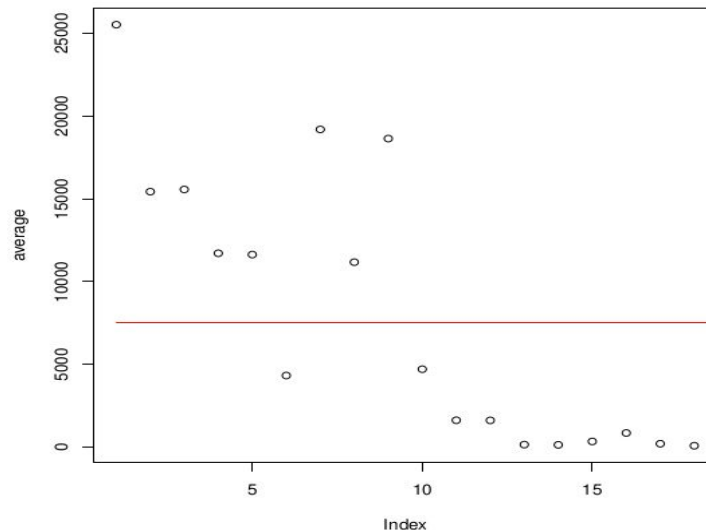
Tarefa 1: identificar padrões entre as marcas de automóvel





Tarefa 1: identificar padrões entre as marcas de automóvel

- Através do gráfico da média de emplacamentos de cada marca é possível identificar uma separação natural, a qual foi utilizada para definir 2 grupos
 - As marcas do grupo com mais emplacamentos (acima da linha vermelha) foram definidas como marcas populares
 - As marcas com valores menores foram definidas como de luxo





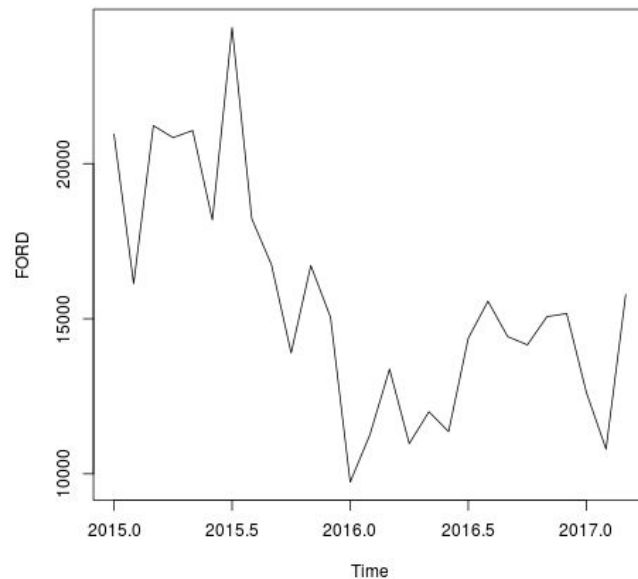
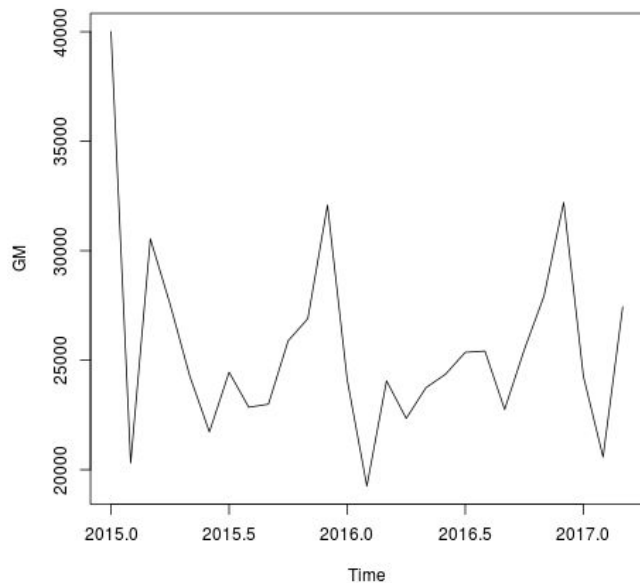
Tarefa 1: identificar padrões entre as marcas de automóvel

- Marcas populares
 - GM, HYUNDAI, FORD, RENAULT, TOYOTA, FIAT, HONDA e VOLKSWAGEN
- Marcas de luxo
 - JEEP, NISSAN, PEUGEOT, CITROEN, KIA, MERCEDES BENZ, AUDI, MITSUBISHI, BMW e LAND ROVER

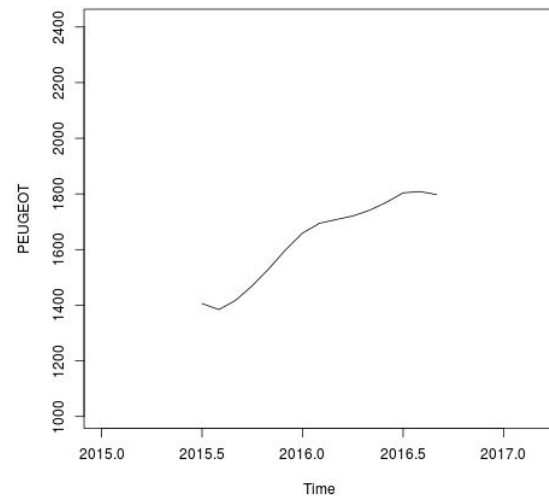
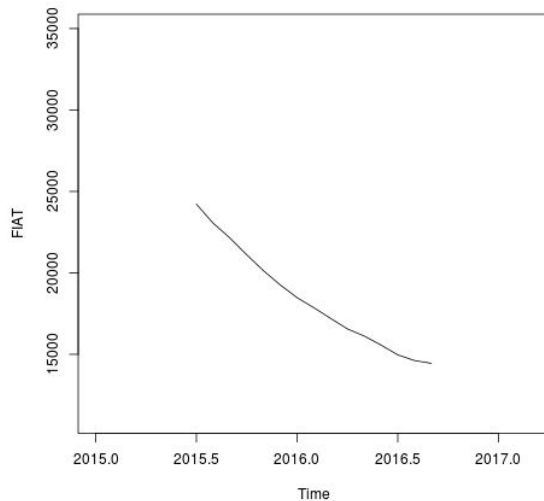
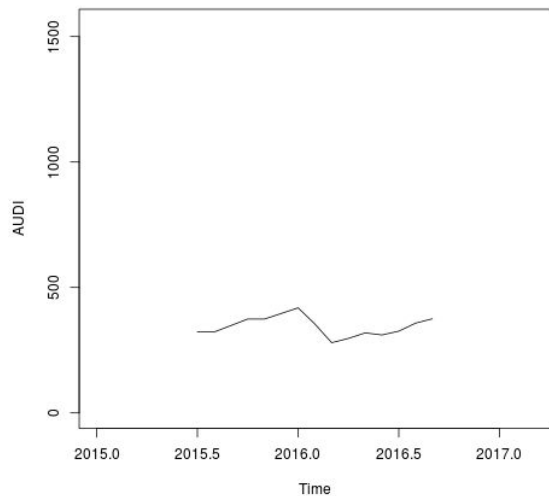


Tarefa 1: identificar padrões entre as marcas de automóvel

- Os gráficos das séries de emplacamentos de cada marca caracterizam o estado atual dessas marcas , porém têm um comportamento com grande variabilidade, tornando-o difícil de avaliar ao observar
- Uma solução é visualizar os gráficos de tendências onde é possível ver o comportamento geral da série e caracterizá-las quanto à direção
- A partir do comportamento da tendência espera-se dividir as marcas entre aquelas que tem um comportamento estável, de crescimento ou queda
 - Para fins de avaliação binária, marcas em crescimento ou estáveis foram denominadas como prósperas e marcas em queda foram denominadas não-prósperas



Exemplos de séries das marcas GM e FORD (todas as outras séries podem ser encontradas em <https://github.com/vfpimenta/auto-project/tree/master/plots/>)



Exemplos de tendências das marcas: AUDI, considerada estável; FIAT, considerada em queda; e PEUGEOT, considerada em crescimento (outros gráficos de tendência podem ser encontrados em <https://github.com/vfpimenta/auto-project/tree/master/plots/>)

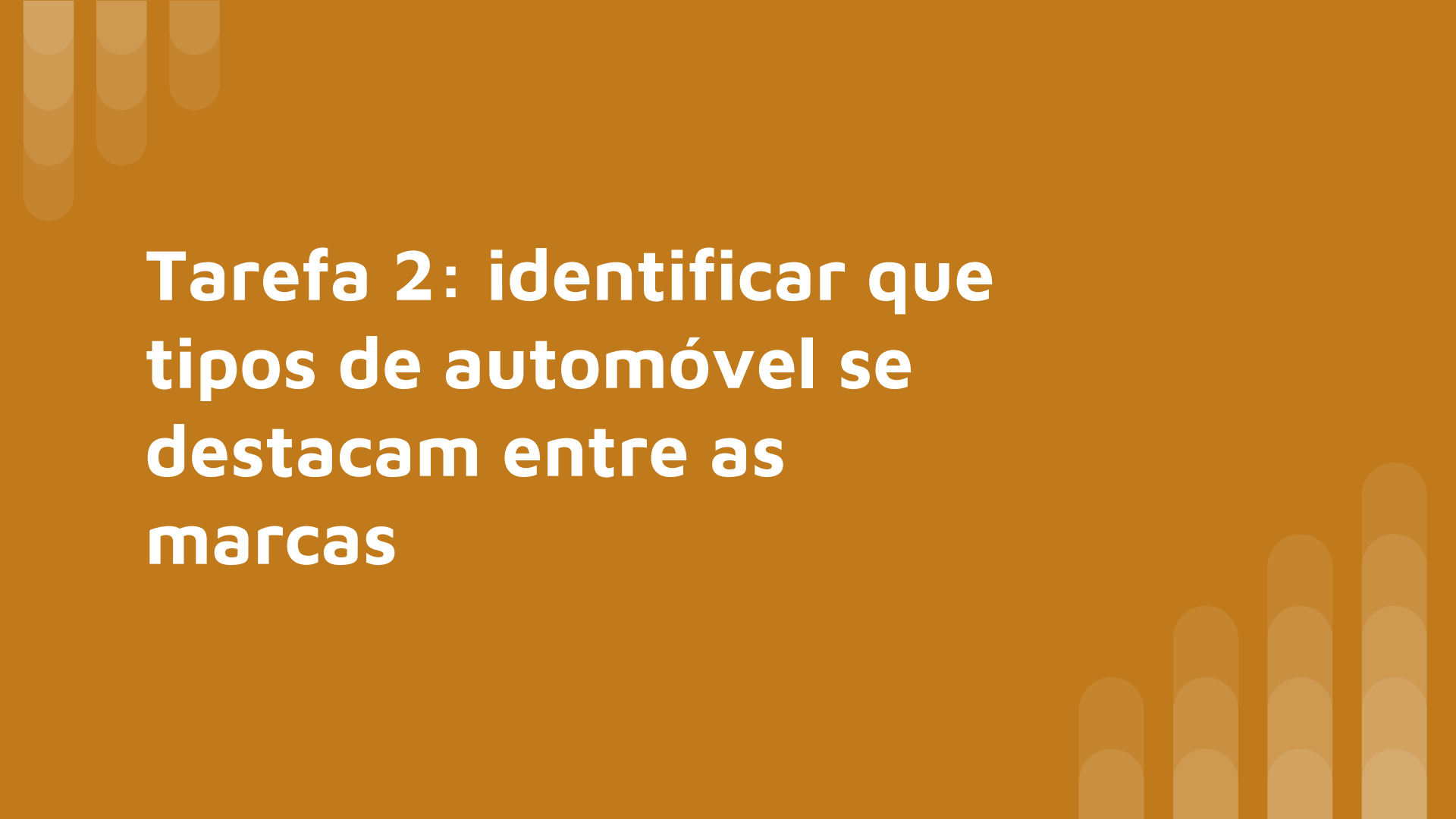


Tarefa 1: identificar padrões entre as marcas de automóvel

- Marcas populares
 - Prósperas
 - GM, HYUNDAI e TOYOTA
 - Não-prósperas
 - FORD, RENAULT, FIAT, HONDA e VOLKSWAGEN
- Marcas de luxo
 - Prósperas
 - JEEP, NISSAN, PEUGEOT, CITROEN, KIA, MERCEDES BENZ, AUDI, BMW e LAND ROVER
 - Não-prósperas
 - MITSUBISHI

Conclusão:

- Marcas de luxo são mais propensas a prosperarem, apesar de venderem menos automóveis
- Marcas populares costumam ter uma tendência mais incerta, apesar de terem mais emplacamentos



**Tarefa 2: identificar que
tipos de automóvel se
destacam entre as
marcas**



Tarefa 2: identificar que tipos de automóvel se destacam entre as marcas

- Tipos de automóveis produzidos por marcas populares
 - HATCH PEQUENO, SEDAN PEQUENO, GRANDCAB, SEDAN COMPACTO, SEDAN MEDIO, SUV, HATCH MEDIO, VEICULO DE ENTRADA, SEDAN GRANDE, SW MEDIO e MONOCAB
- Tipos de automóveis produzidos por marcas de luxo
 - SUV, SEDAN PEQUENO, HATCH PEQUENO, SEDAN MEDIO, MONOCAB e SEDAN GRANDE
- Tipos de automóveis produzidos exclusivamente por marcas populares
 - GRANDCAB, SEDAN COMPACTO, HATCH MEDIO, VEICULO DE ENTRADA e SW MEDIO
- Nenhum tipo de automóvel é produzido unicamente por marcas de luxo



Tarefa 2: identificar que tipos de automóvel se destacam entre as marcas

- Tipos de automóveis produzidos por marcas prósperas
 - HATCH PEQUENO, SEDAN PEQUENO, GRANDCAB, SEDAN COMPACTO, SEDAN MEDIO, SUV, HATCH MEDIO, VEICULO DE ENTRADA, MONOCAB e SEDAN GRANDE
- Tipos de automóveis produzidos por marcas não-prósperas
 - VEICULO DE ENTRADA, SUV, SEDAN PEQUENO, HATCH PEQUENO, SEDAN MEDIO, HATCH MEDIO, SEDAN GRANDE, SEDAN COMPACTO, SW MEDIO e MONOCAB
- Tipos de automóveis produzidos exclusivamente por marcas prósperas
 - GRANDCAB
- Tipos de automóveis produzidos exclusivamente por marcas não-prósperas
 - SW MEDIO

Conclusão:

- **Marcas de luxo são mais seletivas quanto ao tipo de automóvel produzido, enquanto marcas populares possuem um catálogo mais vasto**
- **Dois tipos de automóveis (GRANDCAB e SW MEDIO) se destacaram quanto à prosperidade e devem ser investigados**



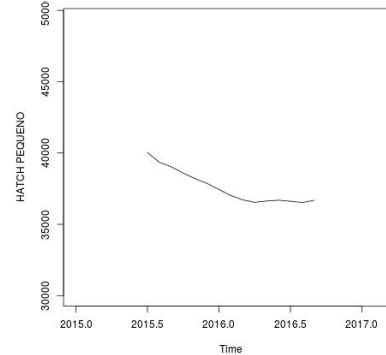
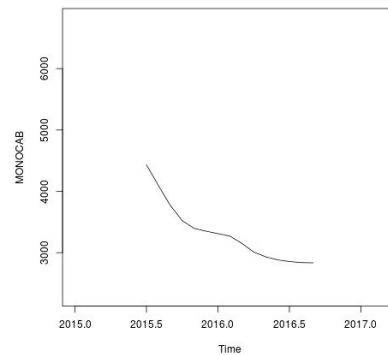
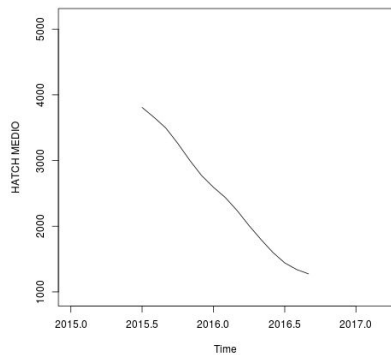
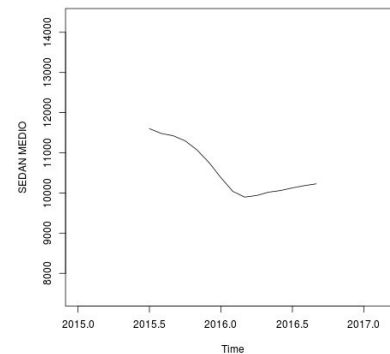
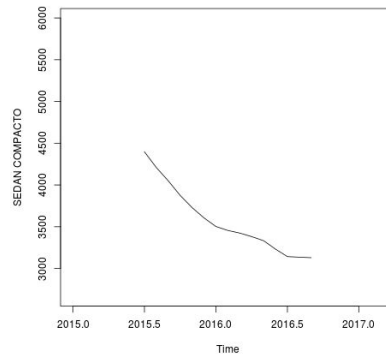
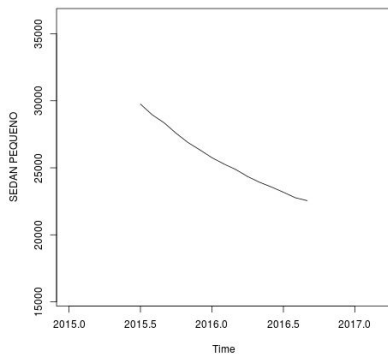
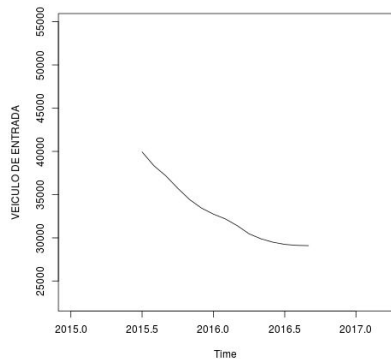
Tarefa 3: identificar o comportamento padrão dos tipos de automóveis



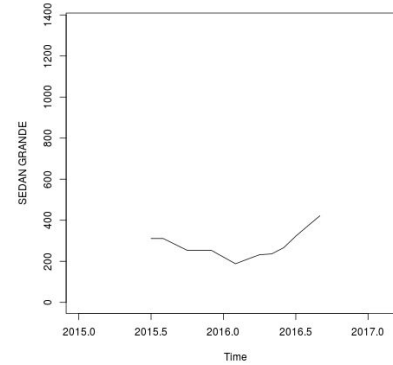
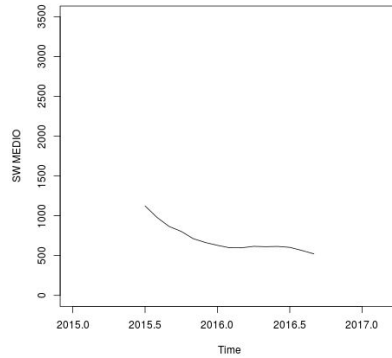
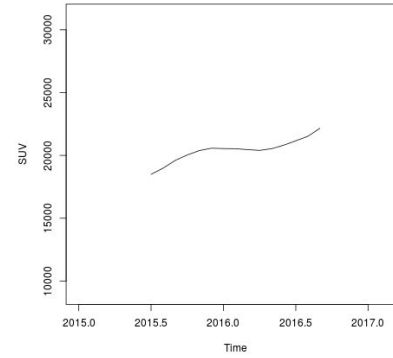
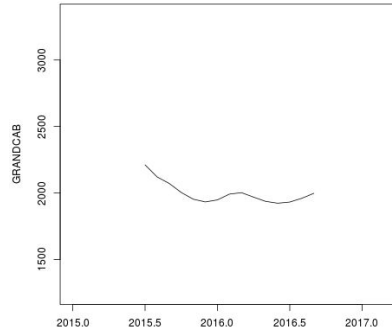


Tarefa 3: identificar o comportamento padrão dos tipos de automóveis

- Através dos resultados da tarefa 2 foi possível identificar alguns tipos de interesse, porém antes de traçar conclusões é importante analisar o comportamento desses tipos
- Observando os valores das tendências dos tipos é possível avaliar se os tipos exclusivos das marcas prósperas e não-prósperas são responsáveis pelo comportamento obtido da marca
- É possível identificar que a maior parte dos tipos de automóvel possuem uma curva de tendência fortemente decrescente, isso é explicado pelo fato que temos várias marcas populares (com grandes números) que são não-prósperas como visto através da tarefa 1, o que leva as curvas a tenderem a um comportamento negativo
- Tipos com comportamentos positivos
 - SEDAN GRANDE e SUV



Exemplos de tendências dos tipos de automóvel. É possível ver que há um padrão típico de decrescimento das séries.



Na esquerda podemos ver os comportamentos dos tipos de interesse GRANDCAB E SW MEDIO. Na direita podemos ver gráficos de tendência crescentes em SUV e SEDAN GRANDE, uma exceção à média.



Tarefa 3: identificar o comportamento padrão dos tipos de automóveis

- Quanto aos tipos de interesse podemos ver que o GRANDCAB tem comportamento decrescente, porém menos acentuado do que outros tipos, o que pode indicar uma vantagem do modelo
 - GRANDCAB é produzido exclusivamente pela GM, uma marca popular próspera, o que levanta dúvidas sobre a validade da sua projeção
 - SW MEDIO é produzido exclusivamente pela VOLKSWAGEN, uma marca popular não-próspera, o que levanta dúvidas sobre a validade da sua projeção
- Quanto ao SEDAN GRANDE há dados faltantes em várias datas, o que lhe desqualifica de análises mais profundas
- Finalmente, o SUV é um tipo de veículo produzido por quase todas as marcas analisadas (exceto VOLKSWAGEN e FIAT) e possui uma tendência fortemente positiva

Conclusão:

- Há evidências de que o GRANDCAB pode ser um tipo de automóvel rentável para outras marcas
- O SUV é o tipo de automóvel que representa o investimento mais sólido entre todos os analisados

Conclusões Gerais

- Apesar das séries disponíveis serem pequenas é possível identificar padrões e pontos de interesse entre os dados analisados
- Pôde-se traçar uma relação entre o fator popularidade e prosperidade
- Pôde-se identificar dois tipos de automóveis com comportamento interessantes que podem ser mais explorados através de dados mais extensos e análises mais profundas