**Análises de Frequência**

* **Frequência absoluta**: quantas vezes cada número (1-50) e cada trevo (1-6) saíram
* **Frequência relativa**: percentual de participação de cada número
* **Números mais e menos sorteados** (quentes e frios)
* **Análise temporal da frequência**: como a frequência muda ao longo do tempo

**Análises de Padrões e Sequências**

* **Números consecutivos**: quantas vezes saíram números em sequência (ex: 15, 16, 17)
* **Repetições entre concursos**: números que se repetem em sorteios consecutivos
* **Intervalos de ausência**: há quanto tempo cada número não sai
* **Ciclos de retorno**: tempo médio para um número voltar a ser sorteado

**Análises de Distribuição**

* **Paridade**: proporção de números pares vs ímpares por concurso
* **Distribuição por dezenas**: quantos números saem em cada faixa (1-10, 11-20, etc.)
* **Soma dos números**: valor total dos 6 números sorteados
* **Amplitude**: diferença entre o maior e menor número do concurso

**Análises de Combinações**

* **Duplas, ternas, quadras**: combinações que mais se repetem
* **Números que saem juntos**: quais têm maior afinidade
* **Padrões geométricos**: distribuição no volante (cantos, bordas, centro)
* **Sequências aritméticas**: números em progressão aritmética

**Análises dos Trevos da Sorte**

* **Frequência dos trevos** (1-6)
* **Combinações de trevos**: quais duplas mais saem
* **Correlação**: relação entre trevos e números principais

**Análises Temporais**

* **Tendências sazonais**: padrões por mês/período do ano
* **Análise de séries temporais**: evolução dos padrões ao longo do tempo
* **Períodos de "seca"**: números que ficam muito tempo sem sair

**Análises Estatísticas Avançadas**

* **Desvio padrão**: variabilidade na distribuição
* **Teste de aleatoriedade**: verificar se os sorteios são realmente aleatórios
* **Análise de clusters**: agrupamentos de números
* **Correlação entre números**: quais tendem a sair juntos
* **Probabilidades condicionais**: chance de um número sair dado que outro saiu

IA GEMINI

**Insights Iniciais (Após rodar a EDA):**

Ao rodar o script de EDA com os dados da Mais Milionária, você provavelmente observará os seguintes tipos de padrões e discrepâncias:

**1. Frequência de Sorteio:**

* **Números Principais (1-50):** Alguns números aparecerão com frequência significativamente maior ou menor que outros. Isso é normal, mas números muito abaixo da média podem ser "tendência de não sair" ou "atrasados", e os muito acima, "quentes".
  + *Discrepância:* Quaisquer números com frequência 0 (se o histórico for curto) ou muito próxima de 0.
* **Trevos (1-6):** Dada a pequena quantidade de trevos, a distribuição de frequência tende a ser mais uniforme, mas ainda pode haver um ou outro trevo ligeiramente mais ou menos sorteado.

**2. Repetições do Concurso Anterior:**

* **Números Principais:**
  + A Mais Milionária sorteia 6 números de 50. A probabilidade de repetição é diferente da Lotofácil (15 de 25). Espera-se um número médio de repetições menor.
  + **Padrão:** Pode haver uma moda (o número mais comum de repetições) e uma faixa esperada. Por exemplo, talvez 1 ou 2 repetições sejam o mais comum. Raramente 0 ou 3+ repetições.
  + *Discrepância:* Concursos com 0 repetições ou um número muito alto de repetições (ex: 4-5) seriam atípicos.
* **Trevos:**
  + Com apenas 2 trevos de 6, a repetição é mais provável. Pode ser comum ter 0, 1 ou 2 trevos repetidos.

**3. Distribuição Pares/Ímpares:**

* **Números Principais:** Dos 6 números, a distribuição mais comum tende a ser próxima de 3 pares e 3 ímpares, ou 2/4, 4/2. É improvável ter 6 pares ou 6 ímpares.
  + **Padrão:** A moda será a proporção mais balanceada (e.g., 3-3 ou 2-4/4-2).
  + *Discrepância:* Concursos com desequilíbrio extremo (e.g., 6 pares, 0 ímpares) seriam outliers.
* **Trevos:** Dos 2 trevos, pode ser 1 par/1 ímpar, 2 pares ou 2 ímpares.

**4. Distribuição de Primos:**

* **Números Principais:** Dos 6 números sorteados de 50, e com 15 primos até 47, o número de primos tende a ser 1, 2 ou 3.
  + **Padrão:** A moda será a contagem mais frequente de números primos.
  + *Discrepância:* 0 primos ou 4+ primos podem ser menos comuns.
* **Trevos:** Os primos entre 1-6 são 2, 3, 5. Então, para os 2 trevos, o número de primos pode ser 0, 1 ou 2.

**5. Soma dos Números:**

* **Números Principais:** A soma dos 6 números terá uma faixa esperada. Uma distribuição gaussiana (curva em sino) é provável, com a maioria das somas caindo em torno da média.
  + *Discrepância:* Somas muito baixas ou muito altas seriam raras e podem indicar um palpite "fora da curva".
* **Trevos:** A soma dos 2 trevos terá uma faixa bem menor (min 1+2=3, max 5+6=11), também com uma distribuição típica.

**6. Sequências Consecutivas:**

* **Padrão:** Com 6 números de 50, é menos provável ter sequências longas (3+ números) do que na Lotofácil. A maioria dos concursos pode não ter sequências, ou ter apenas uma sequência de 2 números.
  + *Discrepância:* Ter uma sequência de 3 ou mais números seria raro.

**Como esses insights ajudam a montar o script:**

* **Definição de Ranges de Filtro:** As modas e as faixas interquartis (25% a 75%) das distribuições (repetições, pares/ímpares, primos, soma, sequências) serão seus novos "limites de ouro" para a função validar\_numeros\_avancado\_final e para a gerar\_combinacoes\_com\_filtros da Mais Milionária.
* **Features para o ML:** As mesmas categorias de features (repetições, pares, primos, soma) serão os *targets* do seu modelo de ML para a Mais Milionária. A complexidade aumenta porque você terá que decidir se o ML prediz:
  + Repetições (números), Pares (números), Primos (números), Soma (números)
  + Repetições (trevos), Pares (trevos), Primos (trevos), Soma (trevos)
  + Ou uma combinação/agregação dessas.
* **Estratégias de Palpites:** Se um padrão é muito forte (ex: "sempre 1 ou 2 repetições de números principais"), você pode dar um peso maior a esse filtro na geração de palpites.
* **Anomalias:** Se o ML predizer algo que está muito fora dos padrões históricos (detectado pela EDA), isso pode ser um sinal para o sistema adaptativo ajustar seus hiperparâmetros ou buscar novas tendências.

**Sugestões/Considerações Adicionais (para complementar o que já está ótimo):**

1. **Visualização como Saída da EDA:** Para cada uma dessas análises, pense em como você pode visualizá-las. Gráficos de barras para frequência, gráficos de linha para tendências temporais, heatmaps para correlações de duplas/ternas, gráficos de dispersão para soma/amplitude, etc. A visualização é a alma da EDA.
2. **Definição de "Quentes" e "Frios":** Para os números "quentes e frios", defina claramente o critério. É pela frequência total? Pela frequência nos últimos N concursos? Pelo intervalo de ausência? Ter um critério claro ajuda na interpretação.
3. **Normalização/Escalonamento (para ML):** Ao transformar essas análises em features para um modelo de ML, lembre-se da importância de normalizar ou escalonar os dados (por exemplo, usando StandardScaler ou MinMaxScaler) para que features com diferentes escalas não dominem o treinamento.
4. **Features para os Trevos:** Para os trevos, além da frequência e combinações, considere também:
   * **Soma dos Trevos:** A soma dos dois trevos.
   * **Produto dos Trevos:** O produto dos dois trevos.
   * **Paridade dos Trevos:** Quantos são pares/ímpares.
   * **Proximidade dos Trevos:** A diferença absoluta entre eles.
5. **Análise de "Moldura" e "Centro" para +Milionária:** Assim como na Lotofácil, você pode definir uma "moldura" e um "centro" para os 50 números e analisar quantos números caem em cada região. Isso pode ser um padrão geométrico interessante.
6. **Número de Primos na +Milionária:** Na +Milionária, o conjunto de números é maior (1-50). Calcular a quantidade de números primos em cada sorteio é outra feature a se considerar, assim como você faz na Lotofácil.
7. **Somas dos Números (Refinamento):** Para a "Soma dos números", você pode ir além da soma total e analisar a distribuição dessa soma. Qual é o intervalo mais comum? Quais somas são raras?

**Estrutura dos Módulos Python (como você já tem na Lotofácil):**

Para a +Milionária, sugiro que você mantenha a mesma modularização, o que facilita muito a organização e reusabilidade do código:

* **carregar\_dados\_milionaria.py**: Para carregar e transformar o CSV da +Milionária em uma matriz binária de 50 colunas para as bolas e 6 para os trevos (ou separado, dependendo de como você modelará).
* **calculos\_milionaria.py**: Funções para cada uma das análises que você listou (frequência, pares/ímpares, primos, soma, amplitude, repetições, etc. para as bolas e para os trevos).
* **funcao\_machine\_learning\_milionaria.py / machine\_learning\_evolutivo\_milionaria.py**: Adaptação dos seus modelos de ML para as features da +Milionária. Provavelmente, você terá que prever 6 números principais e 2 trevos, o que pode exigir uma abordagem de múltiplos modelos ou um modelo que lide com multi-output.
* **validacao\_numeros\_preditos\_milionaria.py**: Para validar os palpites gerados, considerando os padrões identificados na EDA da +Milionária.
* **apresentacao\_milionaria.py / graficos\_milionaria.py**: Para exibir os resultados das análises e os palpites de forma clara.

**Conclusão:**

Sua lista é um excelente roteiro para a EDA da +Milionária. Ao implementar cada um desses pontos, você terá um entendimento profundo dos dados e gerará features de alta qualidade para seus modelos de IA. Vá em frente com essa abordagem detalhada!