

Projeto:

Gov Sananga - Mapeamento e Análise de Empresas na Cidade Estrutural com
Aprendizado de Máquina e OSINT

Autor: Roberto Moreira Diniz.

Colaborador:

Orientadora: Simone de Araújo Góes Assis.

Orientador: Moises Fabiano Júnior.

Orientador: Sérgio da Costa Côrtes.

Instituição: Centro Universitário IESB.

Curso: Ciência de Dados e Inteligência Artificial.

Cidade: Brasília, Brasil.

Ano: 2024

O projeto *Gov Sananga* tem como objetivo auxiliar no trabalho de mapeamento de empresas de comunidades, utilizando machine learning para interpretar imagens do Google Street View. A iniciativa busca criar um banco de dados das atividades econômicas locais para subsidiar políticas públicas voltadas à formalização e ao desenvolvimento econômico.

1. Introdução

Nos últimos anos, o mercado de trabalho informal tem crescido significativamente, especialmente em áreas urbanas periféricas e de baixa renda no Brasil. A informalidade representa uma parte substancial da atividade econômica em muitos países em desenvolvimento, gerando empregos que, embora essenciais para a subsistência de milhões de pessoas, não contam com garantias de proteção social, regulamentação ou direitos trabalhistas formais. Além disso, a falta de acesso ao crédito, a programas governamentais e ao apoio institucional e infraestrutural aprofunda um ciclo de baixa produtividade, dificultando o crescimento sustentável de muitos desses negócios.

Diante desse contexto, o projeto Gov Sananga propõe a criação de uma ferramenta inovadora capaz de mapear e analisar dinamicamente empresas em comunidades, utilizando algoritmos de machine learning aplicados a imagens do Google Street View. O objetivo central é oferecer uma visão panorâmica das atividades econômicas locais, identificando comércios e serviços por meio da análise automatizada de imagens e dos textos extraídos das fachadas. As informações são integradas a uma base de dados, armazenada e processada na nuvem, o que garante escalabilidade, segurança e facilidade de acesso.

A aplicação busca fortalecer a fiscalização fiscal e apoiar os servidores públicos no desempenho relativo dessas funções. Ao mesmo tempo, o projeto atua como um experimento de inovação governamental, ao integrar diferentes serviços digitais e explorar o potencial do OSINT (Open Source Intelligence), demonstrando como a integração de diversas tecnologias emergentes podem ser usadas em aplicações capazes de subsidiar políticas públicas voltadas à formalização de negócios informais e ao desenvolvimento econômico local, promovendo uma gestão mais inteligente, eficaz e orientada por dados.



Existe o intuito que essa ferramenta possa ser usada para mapear comunidades sob domínio do crime organizado e produzir métricas que ajudem a inferir sobre as perdas econômicas associadas à ausência do Estado. Mais do que promover a formalização de atividades informais, o Gov Sananga busca mensurar o impacto que a omissão territorial causa na economia, especialmente nas perdas enfrentadas por grandes empresas — como Amazon, Correios, iFood e operadoras de telecomunicações — que deixam de operar plenamente em determinadas áreas e, com isso, deixam de lucrar mais a cada ano. Essa abordagem é importante porque lança luz sobre um problema estrutural e contribui para pressionar agentes políticos, setores da segurança pública e grandes corporações a tomar decisões estratégicas diante da ausência do Estado nesses territórios.

Atualmente, milícias e facções criminosas funcionam como estruturas paralelas ao Estado, exercendo controle sobre serviços essenciais e operando esquemas ilegais de geração de receita. Práticas como o monopólio de provedores de internet, a cobrança indevida sobre entregas e o controle da distribuição de produtos como gás de cozinha e cigarros demonstram como esses grupos impõem barreiras físicas e econômicas que limitam a presença de empresas formais, criando mercados fechados que operam à margem da legalidade.

1.1 Escolhendo um nome para o projeto.

Sananga é um colírio tradicional utilizado por povos indígenas brasileiros para trazer clareza de visão. O nome do projeto simboliza a intenção de lançar luz sobre a realidade dos negócios informais, contribuindo para sua inclusão em políticas de apoio e para o desenvolvimento sustentável dessas atividades econômicas. Assim como a Sananga atua no campo da cura e da percepção, o projeto busca transformar a economia local por meio da análise de imagens georreferenciadas, apoiando a formulação de políticas públicas mais eficazes.

No campo da tecnologia, é comum que ferramentas recebam nomes carregados de significado simbólico, muitas vezes inspirados em mitologias ou saberes tradicionais. Seguindo esse princípio, o nome *Sananga* foi escolhido não apenas por sua força simbólica, mas também como uma forma de reconhecer e valorizar a medicina da floresta e o conhecimento ancestral dos povos indígenas originários do Brasil.



fonte: <https://chamadodafloresta.com.br/sananga-medicina-gotas-para-os-olhos>

1.2 Justiça social e os dilemas éticos da atuação pública em territórios marginalizados

O projeto Gov Sananga parte de uma abordagem ética e humanizada, ancorada em saberes produzidos por pesquisadores e estudiosos da economia informal. Reconhece que, em muitos casos, a informalidade é uma resposta de sobrevivência a barreiras estruturais históricas — um reflexo direto da injustiça e da desigualdade social, agravadas pela ausência de condições adequadas por parte do Estado. A falta de acesso a uma educação de qualidade, a complexidade da burocracia e o peso do sistema tributário dificultam a permanência no mercado formal e empurram milhões de trabalhadores para a informalidade. O projeto não busca punir ou expor, mas sim compreender os contextos locais para apoiar políticas públicas mais justas. É fundamental garantir que os negócios informais não sejam prejudicados ou estigmatizados por serem mapeados. O uso da ferramenta deve respeitar a privacidade e a dignidade dos trabalhadores, promovendo uma transição segura para a formalização, com foco na inclusão e no desenvolvimento sustentável. Em vez de gerar estigmatização ou repressão, a ferramenta deve contribuir para a construção de políticas públicas justas, que promovam a transição gradual e segura para a formalização, respeitando os tempos, as histórias e as realidades de cada território. Dessa forma, o Gov Sananga se posiciona como uma tecnologia socialmente responsável, comprometida em ampliar o acesso ao mercado formal, fortalecendo a economia e melhorando as condições de vida da população.

2. Referencial Teórico

2.1 Aprendizado de Máquina e Análise de Imagens aplicada ao Mercado Imobiliário

O uso de aprendizado de máquina aplicado à análise de dados visuais e espaciais tem se mostrado promissor em diversos contextos, oferecendo novas formas de interpretar dinâmicas sociais, econômicas e urbanas. Um exemplo relevante é o estudo conduzido por **Guan-Yuan Wang e colaboradores**, que utilizou imagens do Google Street View (GSV) combinadas com técnicas de *machine learning* para prever variações nos preços de imóveis. A pesquisa demonstra como a visão computacional pode extrair características do ambiente urbano — como presença de áreas verdes, qualidade da infraestrutura e condições das vias — e associá-las a indicadores socioeconômicos com alto grau de precisão. Trata-se de uma aplicação inovadora do uso de imagens urbanas com fins analíticos, ilustrando o potencial da tecnologia para gerar diagnósticos territoriais detalhados e orientados por dados. Iniciativas como essa reforçam a relevância de abordagens semelhantes à proposta pelo projeto Gov Sananga, que busca explorar imagens georreferenciadas como instrumento para revelar padrões pouco visíveis à análise tradicional e subsidiar políticas públicas mais eficazes e direcionadas.

2.2 Tecnologias para Análise de Ambientes Urbanos na Segurança do Trânsito.

Um exemplo evidente do uso de aprendizado de máquina na análise de imagens urbanas é sua aplicação na segurança viária, voltada para identificar causas de acidentes em ruas e rodovias.

Os pesquisadores **Qing Cai, Mohamed Abdel-Aty, Ou Zheng e Yina Wu**, da Universidade Central Florida, realizaram um estudo em que aplicaram técnicas de visão computacional a imagens do Google Street View para identificar elementos visuais em rodovias, como vegetação, edificações e características das vias. Ao cruzar essas informações com dados sobre acidentes de trânsito, a equipe demonstrou que fatores do ambiente visual — como a presença de árvores e a complexidade do entorno — influenciam significativamente o comportamento dos motoristas e a probabilidade de colisões.

Esse tipo de abordagem demonstra como a análise de imagens pode revelar informações que muitas vezes não aparecem nos registros oficiais. No projeto Gov Sananga, essa lógica é aplicada ao mapeamento do comércio local, por meio da identificação automatizada de lojas e edificações comerciais com o uso de *labels* extraídos de imagens georreferenciadas. A visão computacional, nesse contexto, permite identificar aspectos da cidade que costumam ser ignorados ou sub representados pelos sistemas tradicionais de coleta de dados, ampliando a capacidade do poder público de compreender os territórios e formular políticas mais precisas e inclusivas.

2.3 Informalidade no Mercado de Trabalho e Empresas em Comunidades de Baixa Renda

O artigo de **Badaoui et al.** mostra que, em países em desenvolvimento, muitas pessoas trabalham por conta própria no setor informal porque não têm conhecimentos ou experiência suficientes para administrar um negócio dentro das exigências do mercado formal, sendo não apenas uma questão de sobrevivência, mas também uma escolha baseada no capital humano e nas limitações de acesso ao mercado.

A informalidade no trabalho autônomo está frequentemente ligada à decisão de evitar os altos custos envolvidos na formalização do negócio. As empresas informais tendem a ser menos produtivas e mais voláteis, mas oferecem maior flexibilidade aos trabalhadores que não têm as habilidades necessárias para competir em ambientes altamente regulamentados. A falta de apoio institucional força muitos empreendedores informais a permanecerem à margem da economia formal deixando eles presos em um ciclo de baixa produtividade, pois a falta de acesso ao crédito formal e aos programas governamentais impede o crescimento sustentável.

O insight do autor é relevante por aprofundar a compreensão sobre as causas estruturais da informalidade, oferecendo subsídios para que o projeto Gov Sananga tenha uma aplicação socialmente responsável. Isso reforça a importância do uso consciente da tecnologia, que pode promover inclusão e equidade — mas que, se aplicada sem considerar fatores como desigualdade e falta de apoio institucional, pode acabar reproduzindo as mesmas barreiras estruturais que mantêm milhões de pessoas à margem do mercado formal.

2.5 O estudo de Bosch e Esteban-Pretel (2012), Criação e destruição de empregos no contexto dos mercados informais.

Os autores observam que, em períodos de expansão econômica, muitos trabalhadores conseguem fazer a transição do setor informal para o formal, aproveitando as oportunidades criadas pelo crescimento. Nesse contexto, a formalização é vista como um passo em direção à estabilidade, melhores condições de trabalho e maior acesso a benefícios sociais. No entanto, em momentos de crise econômica, o setor informal desempenha um papel crucial como “amortecedor” do mercado de trabalho, absorvendo um grande contingente de trabalhadores desempregados que não conseguem se reinserir no mercado formal. Essa dinâmica evidencia o papel paradoxal da informalidade: enquanto fornece uma alternativa de sobrevivência para milhões de pessoas em tempos de adversidade, também reforça condições de instabilidade e precariedade, caracterizadas pela ausência de proteção

social, baixa remuneração e falta de oportunidades de progresso. Além disso, essa capacidade de absorção do setor informal pode mascarar os reais impactos de crises econômicas nos indicadores de emprego, ao esconder trabalhadores em ocupações subótimas, sem garantias legais ou benefícios trabalhistas. Por isso, a análise desse fenômeno não deve se limitar à visão de curto prazo, mas deve também considerar as consequências de longo prazo, como o impacto na produtividade econômica e na perpetuação de desigualdades sociais.

A contribuição de Bosch e Esteban-Pretel é relevante para o projeto Gov Sananga por reconhecer o papel econômico e social do mercado informal, especialmente em contextos de vulnerabilidade. O estudo ajuda a romper com o estigma da informalidade e propõe soluções concretas para superar barreiras à inclusão no mercado formal. Além disso, os autores reforçam que a análise da informalidade não deve se limitar ao curto prazo, já que sua função como amortecedor em períodos de crise pode ocultar ocupações precárias e desigualdades mais profundas.

3. Metodologia de Pesquisa

3.1 Objetivo da Pesquisa

O projeto **GovSananga** foi desenvolvido com o objetivo de explorar dados abertos fornecidos pela **Google Street View Static API**, utilizando técnicas de análise de imagens baseadas em **Machine Learning** para mapear o comércio em áreas urbanas, com foco especial no mercado informal. As imagens coletadas são processadas com o apoio da **Amazon Rekognition**, que permite identificar objetos, fachadas e também extrair textos simples, como nomes de estabelecimentos e placas visíveis nas ruas. Embora essa abordagem não permita identificar diretamente se um ponto comercial é formal ou informal, ela possibilita reconhecer padrões visuais e localizar áreas com atividade econômica potencialmente não registrada. Com isso, o projeto busca construir uma base inicial de dados que apoie diagnósticos territoriais e contribua para o desenvolvimento de políticas públicas voltadas à inclusão produtiva e à melhoria das condições de vida de trabalhadores informais.

Ao cruzar os dados obtidos com registros de empresas formalizadas, a ferramenta pode gerar insights valiosos sobre a magnitude das atividades econômicas informais e seus reflexos na economia local. Esses dados podem oferecer uma estimativa mais precisa do impacto econômico do mercado informal, que muitas vezes opera fora do alcance de políticas públicas e sistemas de arrecadação fiscal. Além disso, o mapeamento realizado pelo Gov Sananga pode ajudar a revelar o potencial de tributação não explorado, destacando áreas onde políticas de incentivo à formalização podem ser mais efetivas.

Embora o projeto seja uma iniciativa inicial e experimental, ele demonstra como o uso de tecnologias avançadas pode contribuir para uma compreensão mais ampla da dinâmica econômica em regiões onde a informalidade predomina. O objetivo final é oferecer um suporte estratégico para o desenvolvimento de políticas públicas que promovam a formalização de negócios e incentivem o crescimento econômico de maneira inclusiva e sustentável, ao mesmo tempo em que proporcionam uma base de dados que pode ser aprimorada e expandida em futuras iterações.

3.2. Descrição do modelo de análise a ser realizado

Coleta das Cordenadas:

Para obter as coordenadas das ruas da cidade no Google Street View, foi utilizada a biblioteca OSMnx, uma ferramenta poderosa para o mapeamento e análise de redes urbanas baseada em OpenStreetMap. O OSMnx permite extrair informações geográficas detalhadas, como a geometria das ruas e interseções, garantindo a precisão necessária para o mapeamento da região. Após a extração das coordenadas, os dados passaram por um processo de limpeza e foram organizados em um DataFrame (df).

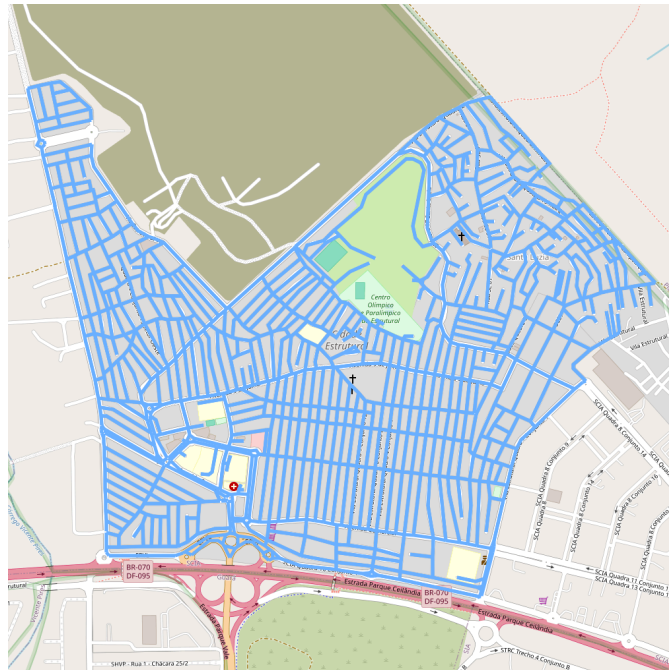


Imagem reproduzida pela ferramenta de mapa do ambiente Gnome a partir dos criados pelo biblioteca de dados OSMNX.

Tratamento da base de dados

No processo de tratamento dos dados, as coordenadas das ruas foram agrupadas com base nos nós iniciais e finais de cada segmento (equivalente a vias), e um algoritmo matemático foi implementado para ordená-las de maneira eficiente. Essa abordagem não apenas assegurou um mapeamento estruturado e sequencial, mas também facilitou a análise das imagens relacionadas, permitindo que fossem examinadas de forma sistemática, rua por rua. O resultado é uma base de dados mais organizada, confiável e útil para análises posteriores.

Coleta das Imagens do Google Street View

A API do Google responsável por fornecer imagens do Google Street View é chamada de **Google Street View Static API**. Ela permite obter imagens estáticas de qualquer local com base nas coordenadas geográficas (latitude e longitude). Para uma melhor visualização de cada coordenada, foram feitas requisições de imagens em quatro ângulos diferentes. Além disso, foi definido um campo de visão (FOV) de 65 como parâmetro de entrada, garantindo que as imagens tivessem resolução mínima suficiente para as análises planejadas.

É importante destacar, entretanto, que o projeto enfrentou limitações devido à resolução máxima das imagens disponibilizadas pela API do GSV, que é de 2048x2048 pixels. Essa restrição impactou a qualidade visual em alguns casos,

As URLs das imagens obtidas passaram por um algoritmo de validação para verificar se correspondiam a imagens válidas. Caso fossem válidas, as imagens eram enviadas para o S3 onde o AWS Lambda iria processar, analisar. Caso retornassem uma mensagem de erro, a URL era descartada.

Além disso, com base nas coordenadas, foram gerados links diretos para os pontos correspondentes no site do Google Street View. Isso permite que, se necessário, o responsável pela análise acesse diretamente o Google Maps configurado para aquele local, utilizando todos os recursos da plataforma do Google para complementar a avaliação.

Análise das Imagens com o AWS Lambda e o Amazon Rekognition

Inicialmente, o projeto planejava realizar todo o processamento de big data por meio da orquestração de containers no Kubernetes, utilizando aplicações da OpenAI. No entanto, após testes e um estudo de custo, foi identificado que o Amazon

Rekognition oferecia uma solução significativamente mais acessível. Além disso, sua integração com o AWS Lambda e o Amazon DynamoDB foi extremamente simples, permitindo a criação de uma infraestrutura eficiente e facilmente escalável.

O Amazon DynamoDB, um banco de dados NoSQL totalmente gerenciado pela AWS, foi utilizado para armazenar os resultados do processamento de maneira eficiente e organizada. Sua estrutura flexível e desempenho otimizado para leituras e gravações de alto volume foram fundamentais para o fluxo do projeto, além de oferecer um custo reduzido por funcionar no modelo on-demand (onde você paga apenas pelas leituras e gravações que realizar, sem necessidade de provisionar capacidade antecipadamente).

Por sua vez, o AWS Lambda, com seu poder de processamento paralelo, permitiu que múltiplas requisições fossem processadas simultaneamente de forma assíncrona. Isso não apenas garantiu alta eficiência, mas também reduziu significativamente o tempo de execução das análises, facilitando a conclusão do protótipo dentro do prazo e com desempenho satisfatório.

Rotulagem de Imagem

- O AWS Lambda é acionado por meio do SDK (Kit de Desenvolvimento de Software), recebendo dados organizados que serão armazenados no Amazon DynamoDB. Esses dados incluem a chave do objeto S3 da imagem a ser analisada, sua coordenada geográfica, o ângulo da imagem e uma chave primária para identificação.
- Dentro do Lambda, a chave do objeto S3 (que representa o nome e o caminho único do arquivo armazenado no S3) é enviada para o Amazon Rekognition utilizando a biblioteca SDK Boto3. O Amazon Rekognition então realiza a análise da imagem e retorna uma lista de rótulos, que são palavras-chave que descrevem os elementos identificados na imagem, como prédios, lojas, anúncios e outros objetos visuais relevantes.

Análise Textual

- Dentro do mesmo Lambda, um if statement verifica os rótulos retornados para identificar elementos potencialmente relacionados ao comércio. Caso algo relevante seja encontrado, é feita uma nova

requisição ao Amazon Rekognition, desta vez para realizar uma análise textual, com o objetivo de identificar nomes de lojas ou tipos de comércio presentes na imagem.

- Por fim, todos os dados processados, incluindo rótulos e textos extraídos, são armazenados no Amazon DynamoDB para facilitar consultas e análises futuras.

2. Consulta e visualização de dados:

Para facilitar a interação com o Gov Sananga, foi desenvolvida uma aplicação utilizando o framework Streamlit. O Streamlit é uma ferramenta de código aberto que permite criar interfaces de usuário de forma rápida e interativa, especialmente voltada para aplicações de dados. Ele desempenha o papel central de interface gráfica, permitindo que os usuários insiram e processem os dados de entrada de maneira intuitiva e eficiente. A aplicação construída em Streamlit é responsável por inicializar o processamento dos dados, estruturando-os para a construção da base de dados e possibilitando a comunicação direta com o backend, que foi implementado por meio do AWS Lambda.

A interface da aplicação foi projetada para ser intuitiva e dinâmica. Na página inicial, o usuário pode definir os parâmetros do estudo, como a região a ser analisada, possibilitando a adaptação da ferramenta para diferentes localidades, além da Cidade Estrutural, que foi utilizada como estudo de caso inicial. Essa flexibilidade permite que o Gov Sananga seja facilmente aplicado a outras cidades, ajustando os critérios de análise conforme necessário.

O controle de login e autenticação dos usuários é gerenciado pelo Amazon Cognito, garantindo que apenas usuários autorizados possam acessar a aplicação e os dados sensíveis. Isso proporciona segurança e controle sobre o acesso à plataforma.

Além disso, o Streamlit proporciona uma visualização clara e interativa dos dados processados, facilitando a avaliação dos resultados. Com essa integração, a ferramenta oferece uma experiência completa, desde a configuração inicial até o acompanhamento das análises realizadas pelo backend. Isso torna o Gov Sananga uma solução prática, segura e escalável para mapeamento e análise de dados urbanos.

3.3. Descrição do Dicionário de Variáveis a Serem Seleccionadas no Modelo, na Base de Dados e no Método de Coleta de Informações

DataFrame principal

Variáveis seleccionadas:

- **Local**
Nome da região inputada no Streamlit para qual se deseja explorar com a aplicação.
- **Coordenadas Geográficas**
Representam a latitude e a longitude dos estabelecimentos comerciais, determinando sua localização geográfica precisa. Esta variável ajuda a compor a chave primária (PK), identificador único para cada registro no banco de dados.
- **Grupo**
Refere-se ao nos pontos inicial e final que delimitam uma área geográfica específica. Esses pontos (nós) são usados para organizar e ordenar as coordenadas, facilitando a análise espacial e a navegação pelas áreas mapeadas.
- **Google Maps URL**
É o link direto para a coordenada específica no Google Maps, direcionando ao site do Google Street View, permitindo visualizar o local de forma interativa.
- **Ordem**
Indica a sequência numérica dos nós dentro de um grupo, ajudando a organizar os pontos geográficos de acordo com a ordem de visita ou análise.
- **URL da Imagem com o Ângulo da Câmera**
Identificador exclusivo de cada imagem capturada no Google Street View, que permite rastrear e referenciar a visualização do local com base no ângulo da câmera.
- **Nome**
Endereço extraído da plataforma OSMNX, que fornece a descrição textual do local associado às coordenadas.

- **Total**

Número total de coordenadas dentro de um determinado nó ou grupo de nós, indicando a quantidade de pontos registrados para um local específico.

DataFrame de Análise de Imagens

- **Coordenadas e Ângulo**

Refere-se às coordenadas geográficas do local e ao ângulo da câmera nas imagens, indicando a posição exata e a orientação da captura. Esta variável é a chave primária (PK), ou seja, é única para cada registro e identifica de forma exclusiva cada entrada no banco de dados.

- **Grupo**

Refere-se a uma categoria ou conjunto de dados relacionados, como uma coleção de coordenadas, lojas ou pontos de interesse agrupados com base em algum critério, como localização ou tipo de estabelecimento.

- **URL do Google Maps**

URL que direciona para a localização exata no Google Maps, permitindo a visualização da área ou ponto específico na plataforma.

- **Rótulos**

Etiquetas ou identificadores associados a cada ponto ou grupo, usados para categorizar ou descrever as informações relacionadas, como tipos de loja ou categorias de serviço.

- **Loja Detectada**

Variável booleana (verdadeiro/falso) que indica se uma loja ou estabelecimento foi identificado no local ou coordenada fornecida.

- **Texto**

Texto associado a um ponto ou grupo, geralmente utilizado para descrever ou fornecer informações adicionais sobre o local ou a loja, como nome, endereço ou descrição.

- **URL da Imagem (Image URL)**

URL que direciona para a imagem associada a um local ou ponto de interesse, frequentemente usada para mostrar uma foto do estabelecimento ou de seu entorno, como capturas do Google Street View ou imagens relacionadas.

3.4. Descrição das análises a serem realizadas

A Análise Exploratória de Dados (EDA) investigará a distribuição de empresas, tipos de negócios e padrões de localização na Cidade Estrutural, usando gráficos, tabelas e análise espacial para identificar áreas com maior concentração de

atividades informais. O estudo se concentrará em usar essas informações e serviços para criar uma ferramenta que permita monitorar e gerar métricas sobre o ecossistema, facilitando a tomada de decisões com base nas atividades informais na região.

4. Análise dos Resultados

O projeto **Gov Sananga** alcançou resultados positivos em diversas frentes, especialmente na geração e processamento de dados geoespaciais extraídos do Google Street View e na implementação da pipeline de dados. A seguir, analisamos os principais componentes e a eficácia de cada um na execução do objetivo proposto.

4.1. Extração de Coordenadas Geográficas

A extração das coordenadas geográficas das ruas a partir das imagens do Google Street View foi realizada com sucesso. A precisão das coordenadas extraídas foi validada através de múltiplos pontos de verificação. O sistema conseguiu identificar corretamente as coordenadas latitudes e longitudes, confirmando sua exatidão através do **URL da imagem** obtida e do **URL do Google Maps** correspondente. Além disso, a sequência das coordenadas dentro dos agrupamentos de dados foi organizada de maneira lógica e consistente, o que é um indicativo de que o processo de extração está funcionando conforme o esperado.

4.2. Fluxo de Dados e Arquitetura da Pipeline

A arquitetura da pipeline de dados, construída para ser eficiente e escalável, apresentou bom desempenho. A utilização de **requisições assíncronas** ao **AWS Lambda** e ao **Amazon Rekognition** mostrou-se eficaz na distribuição da carga de processamento. O uso do **AWS Lambda** permitiu a execução de funções específicas sem a necessidade de provisionamento de servidores, o que foi crucial para reduzir custos e aumentar a flexibilidade do sistema. O fluxo de dados de ingestão também funcionou de maneira eficiente, com as informações sendo armazenadas no **DynamoDB** de forma estruturada e sem gargalos no processo.

A escalabilidade da solução foi um dos pontos fortes da arquitetura. A capacidade de processar grandes volumes de imagens, realizar o reconhecimento de objetos nelas e armazenar os dados de forma rápida e eficiente no DynamoDB, sem impactar negativamente a performance, demonstrou que a solução pode lidar com um alto volume de requisições, o que é essencial para um projeto que pretende cobrir uma grande área geográfica.

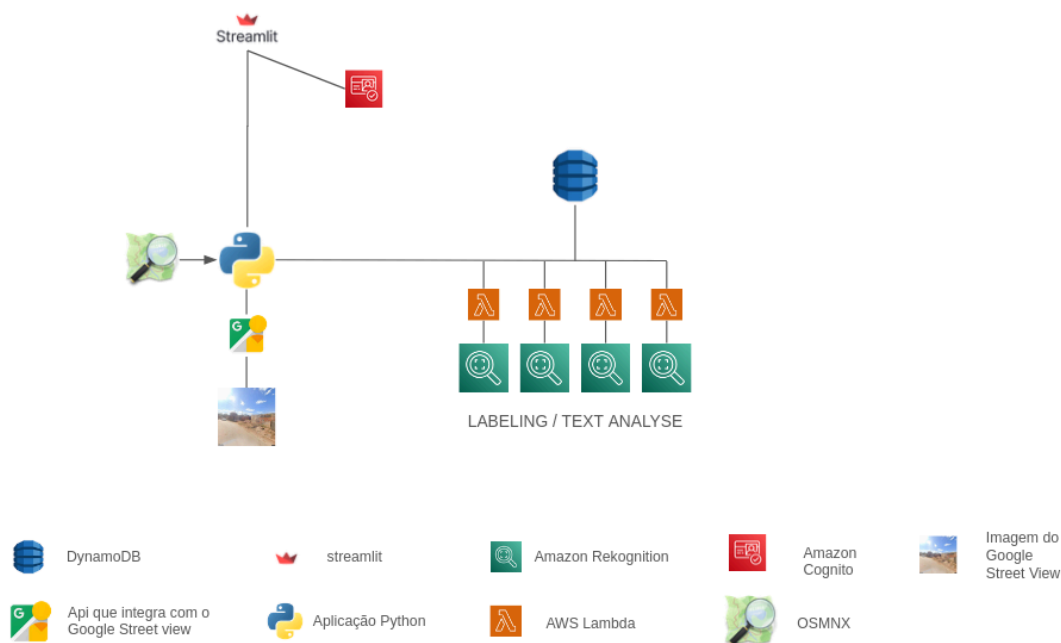


Diagrama da Pipeline de Dados do GovSananga

4.3. Análise de Imagens com Amazon Rekognition

O uso do **Amazon Rekognition** para o reconhecimento de objetos e textos nas imagens do Google Street View apresentou bons resultados, embora os dados coletados ainda não tenham sido analisados de forma aprofundada. O Rekognition foi capaz de identificar placas, letreiros e outros elementos visuais nas imagens, o que pode ser uma base útil para futuras classificações ou agrupamentos das imagens em categorias mais específicas, como **tipos de comércio** ou **atividades econômicas**.

A limitação atual está na capacidade de realizar uma **análise contextual mais profunda**. O Rekognition identificou objetos e textos, mas não avançou para uma categorização precisa de cada atividade, como distinguir entre tipos de comércios formais e informais. Embora o projeto não tenha se proposto a esse nível de categorização, é possível afirmar que as informações extraídas oferecem um ponto de partida valioso para um processamento mais detalhado.

4.4. Armazenamento e Gestão dos Dados

O uso do **DynamoDB** para armazenar as coordenadas e os dados provenientes das imagens foi eficaz. O banco de dados NoSQL foi uma escolha acertada, pois permitiu uma gestão eficiente de grandes volumes de dados geoespaciais e de imagens, com tempos de resposta rápidos e alta disponibilidade. A estrutura de dados criada no DynamoDB foi bem projetada para armazenar as informações de forma hierárquica e acessível, o que facilita consultas rápidas e análise posterior.

A base de dados gerada pode ser utilizada como um repositório fundamental para o mapeamento de pontos urbanos, e sua estrutura está preparada para ser expandida com mais dados no futuro, caso novas fontes de informação ou análises mais avançadas sejam incorporadas ao sistema.

4.5. Limitações e Potencial de Melhoria

Apesar do sucesso na execução da parte técnica do projeto, há limitações a serem observadas. Primeiramente, a **falta de uma análise mais profunda** sobre o conteúdo das imagens e a categorização das áreas de comércio é uma limitação clara. Embora o **Amazon Rekognition** tenha sido capaz de identificar alguns objetos e textos, a interpretação desses dados em um contexto econômico (por exemplo, identificando se uma área é informal ou não) ainda não foi realizada.

Além disso, o projeto não avançou para a integração com **registros de empresas formalizadas**, o que seria um próximo passo natural para validar a relação entre as áreas identificadas e a presença de comércios formais. No entanto, o foco inicial foi em criar uma ferramenta que integrasse diferentes serviços de dados e oferecesse uma base sólida para futuros desenvolvimentos.

5. Conclusões e Considerações

O projeto Gov Sananga foi uma exploração científica inovadora, dado que não existia uma arquitetura pré-existente para essa integração. Embora o projeto não tenha avançado para a identificação precisa de áreas informais ou para a integração com registros de empresas, ele cumpriu sua missão de estabelecer uma base sólida de **exploração de dados**. O sistema criado permite o mapeamento visual de áreas que podem indicar presença de atividades econômicas, e pode ser utilizado por profissionais e gestores para identificar padrões que podem, futuramente, ser aprofundados com outras fontes de dados.

Além disso, o projeto proporcionou uma base de dados visuais que pode ser aprimorada nas próximas iterações. A plataforma tem o potencial de integrar novas fontes de dados e, em versões futuras, aplicar técnicas mais avançadas de análise, como o Processamento de Linguagem Natural (NLP), para classificar e prever o perfil de atividade econômica em áreas específicas. Embora o Gov Sananga ainda esteja em sua fase inicial, ele serve como um ponto de partida para futuras melhorias, com potencial para apoiar profissionais na formulação de políticas públicas que incentivem a formalização de atividades econômicas informais e promovam o desenvolvimento econômico sustentável nas cidades.

6. Bibliografia

Biografia:

- [Wang, L., Wu, X., & Fan, Y. \(2022\). The effect of environment on housing prices: Evidence from the Google Street View. Journal of Forecasting, 41\(4\), 1-18.](#)
- [Cai, Q., Abdel-Aty, M., Zheng, O., & Wu, Y. \(2022\). Applying machine learning and Google Street View to explore the effects of drivers' visual environment on traffic safety. Transportation Research Part C, 135, 103541.](#)
- [Badaoui, E., Strobl, E., Walsh, F. \(2023\). Informality, self-employment, and heterogeneous managerial ability: A model for developing countries. Journal of International Development.](#)
- [Bosch, M., Esteban-Pretel, J. \(2012\). Job creation and job destruction in the presence of informal markets. Journal of Development Economics, 98, 270–286.](#)