



CASE LGPD

DATA LAND

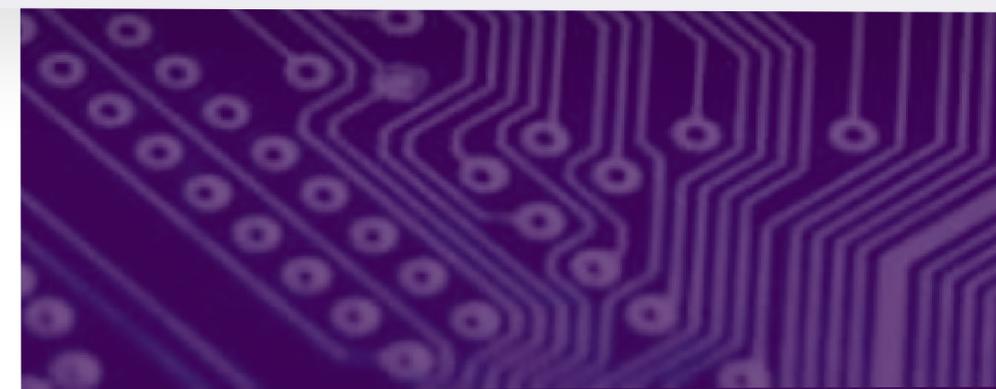
 **DATALAND**
LAND POWERED BY DATA SCIENCE

ORIGEM E O QUE É

a solução DataLand

O setor imobiliário enfrenta diversos desafios de aceleração e crescimento, principalmente por fatores como: longos ciclos de lançamento de imóveis, tributação desfavorável à técnicas mais modernas de construção como a construção modular ou que inviabilize a industrialização seriada de moradias por meio da pré-fabricação. Além da falta de oferta de produtos financeiros de financiamento ou investimento para os clientes finais. Ainda por cima, o setor da construção é o segundo pior em digitalização no mundo inteiro, segundo estudos da McKinsey recentes mostram.

Para mudar este cenário, nós acreditamos no conceito open innovation, que traz uma nova cultura de expansão e a conexão entre os diversos agentes que movem a indústria. A inovação aberta se baseia na difusão de ideias e conhecimentos que acontece, principalmente, a partir de parcerias externas. O fato de se conectar com aquilo que é de fora e de poder compartilhar novas soluções gera um tipo de trabalho colaborativo entre diferentes empresas, startups e iniciativas que possam agregar de alguma forma o setor em questão.



Neste ponto, é fundamental o papel das Construtechs, Proptechs e dos hubs de inovação, pois, estes têm sido uma alternativa para trazer ao ambiente da construção novas alternativas e a integração entre empresas de diferentes nichos, interessados na transformação e no amadurecimento no mercado da construção.

De acordo com o mapeamento de 2020 da Terracotta Ventures, o Brasil identifica 702 Construtechs e Proptechs dos mais diversos nichos da construção civil que estão ativas, apresentando um aumento de 23% em relação à 2019. O mapa também mostra que o mercado de startups para construção civil está dividido em quatro grandes blocos.

OS BLOCOS SÃO:



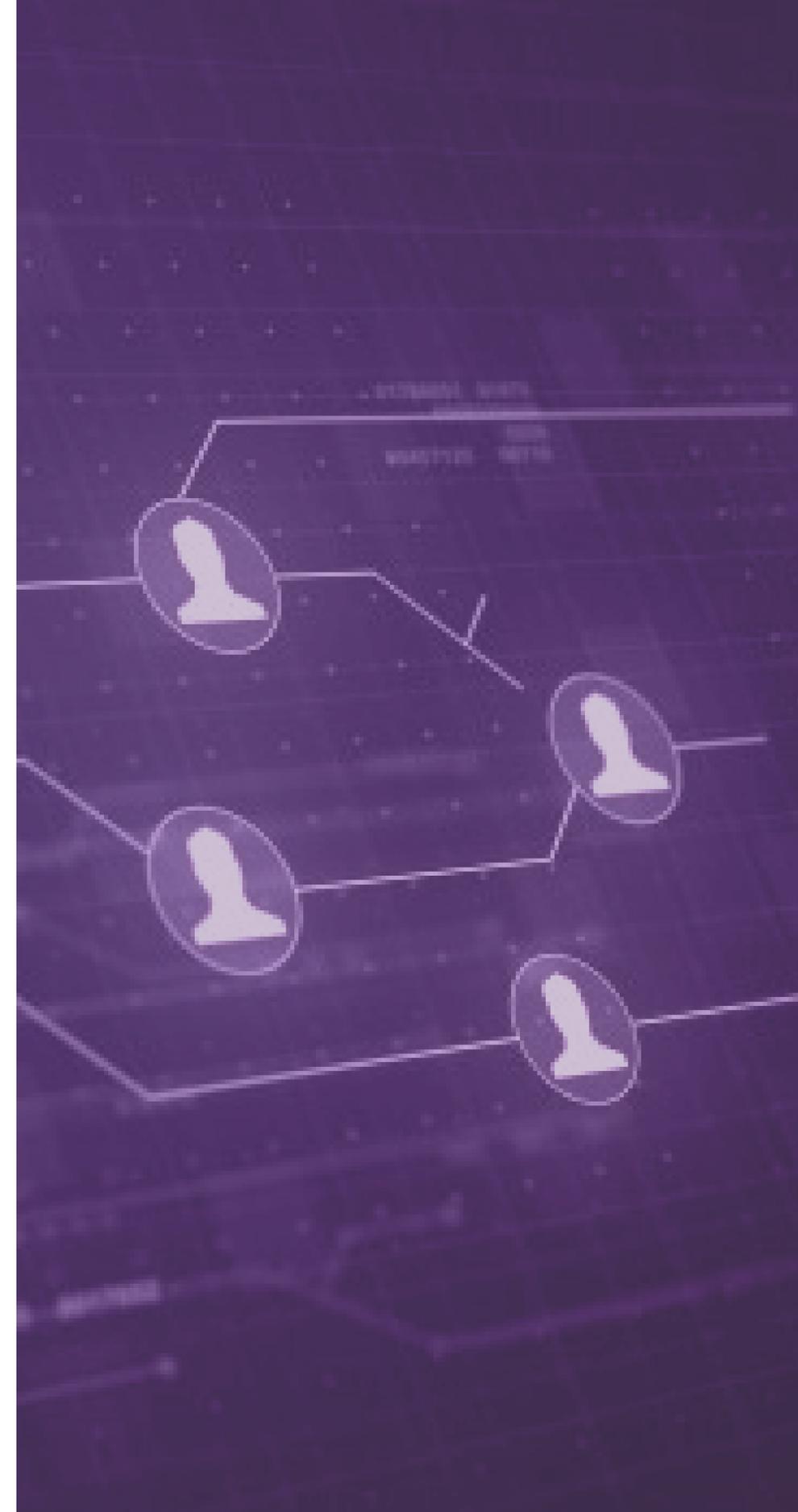
No início de 2020 a DataLand realizou mais de 40 entrevistas de profundidade com incorporadoras e fundos imobiliários para melhor entendimento na fase de projeto e viabilidade, quais eram as reais dores. Através das entrevistas constatamos que o processo de viabilidade de terreno e estudo vocacional, ainda é muito manual, volumoso e intuitivo.

FRASES COMO:

“Atualmente compramos terrenos por intermédio de corretores de terreno que ofertam o terreno ao mesmo tempo para diversos players, gerando um grande volume de análises, com efetividade de 5 a 10% dos casos”.

“As análises atuais utilizam ferramentas proprietárias, em um processo que conta com intuição e informações limitadas. O mesmo é manual, demandando muito tempo para orientar a tomada de decisão e também não permite atualizações automática, fazendo com que os analistas tenham que refazer como se fosse um novo projeto” foram faladas diversas vezes.

Durante as entrevistas também identificamos dois principais perfis de “público”, os que utilizam os dados disponíveis, mesmo que restritos e de diversas fontes, para embasar a tomada de decisão e os que entendem que a experiência e intuição são fundamentais no momento da compra do terreno, com baixa utilização de dados para embasamento da decisão.



Com base nas entrevistas identificamos grande oportunidade e potencial de mercado, pois as dores citadas inspiraram a equipe DataLand na confecção da plataforma **“Urban Data Science”**, pois:

“Uma compra errada de um terreno que chega a valer 15% do VGV, uma incorporação com produto que não é vocação do local, impacta significativamente nos resultados! Além da imagem da empresa que fica abalada no mercado!”

Outro ponto comentado, que nos impactou, foi em relação ao crescimento

das cidades:

“As mesmas crescem de forma desordenada, sem análises profundas de dados estruturados e combinados, pressionando a gestão urbana a uma demanda adicional por infraestrutura”

A DataLand nasce em abril de 2020, com cultura empreendedora, apaixonada e experiente para resolver dores na fase de incorporação, através de dados estruturados, data science e tecnologia disruptiva, de forma prática e acessível.

Como propósito temos

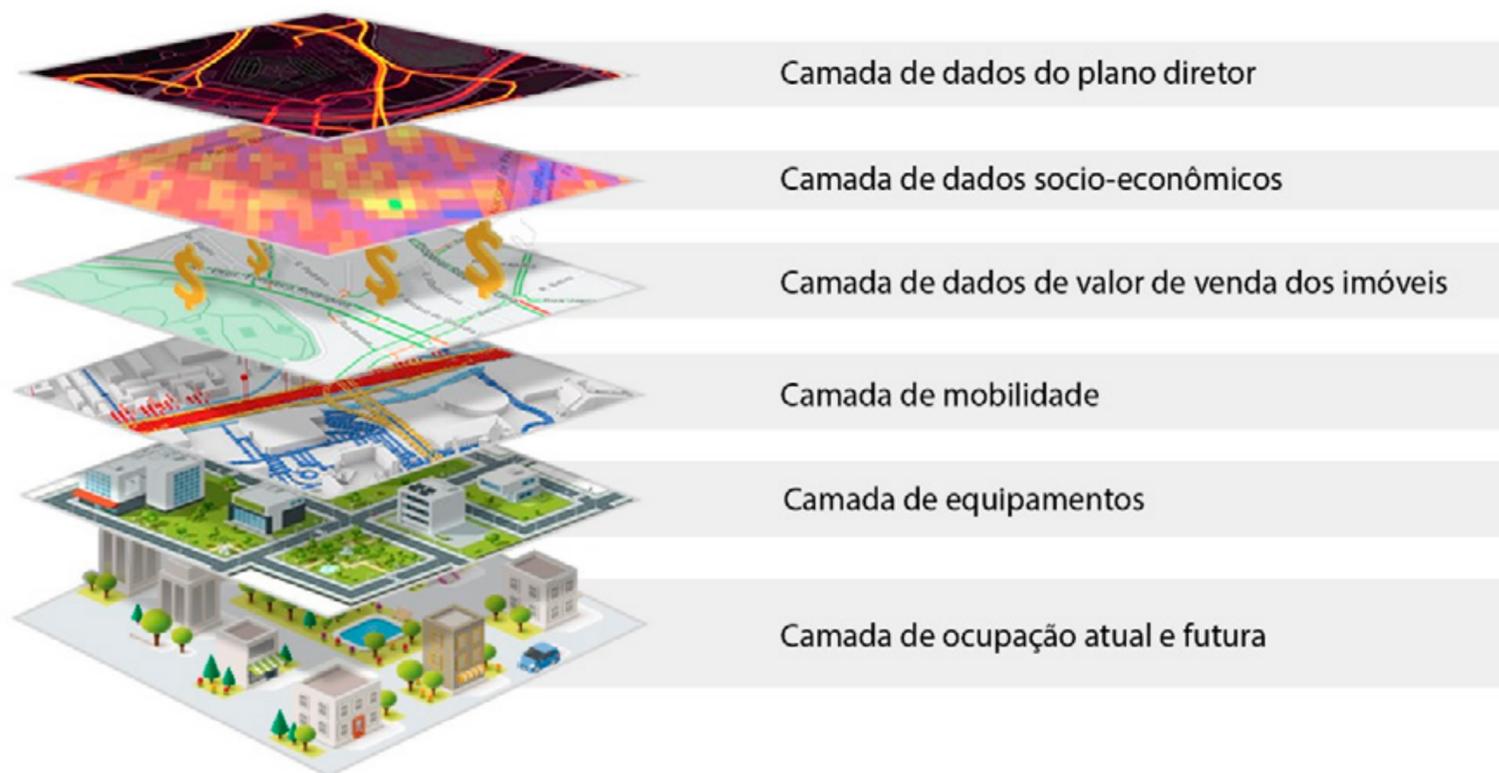
***Transformar a utilização das terras
através de data science***



Em países da Europa, por exemplo, diversos institutos fazem a coleta e disponibilização de dados urbanos, a fim de ajudar legisladores e desenvolvedores imobiliários a desenvolverem projetos melhores para as cidades.

Infelizmente, aqui no Brasil essa prática ainda não está formalizada e coloca um grande desafio para as tomadas de decisão data driven. É por isso que a Data Land se colocou esse desafio de aplicar as metodologias do Big Data e Data Science para abrir o caminho para que as empresas que desenvolvem o mercado imobiliário no Brasil possam tirar proveito das informações que podem estar disponíveis para diminuir os riscos sobre os investimentos e fazer decisões sustentáveis para as cidades no longo prazo.

A Data Land coleta, cataloga e clusterizar dados, estes são combinados para gerar informações, estas analisadas gerando insights para os nossos clientes. Para exemplificar, segue modelo da estruturação da informação DataLand.



Modelo da estruturação da informação DataLand.

Dados clusterizados, gerando informação:

Nós processamos os 3,5 milhões de lotes do município de São Paulo através de mais de 200 camadas de informação.

Conseguimos, de forma ágil, informar os restaurantes, shopping centers, padarias, pet shops, escolas, modais de mobilidade, entre outros aspectos no raio de atuação que fizer sentido ao nosso cliente, bem como, o perfil de empresas que atuam na região.

2 Informações analisadas gerando insights,

conseguimos com grande agilidade informar o tipo de uso e os parâmetros de ocupação, pois estudamos o plano diretor, operação urbana e fatores de incentivo do município de São Paulo, e através da calibragem dos algoritmos em questão de segundos, conseguimos gerar o coeficiente de aproveitamento máximo do terreno, considerando todos os incentivos, o tipo de empreendimento que faz mais sentido na região de acordo com a absorção e perfil de moradores e por fim, a margem operacional.

No nosso modelo, pensamos em agregar fluxo e mobilidade das pessoas nas cidades, com o intuito de proporcionar as melhores opções de fachada ativa e comércio de acordo com o perfil da região.





Falando em visão futura, a nossa proposta é elaborar análises integradas e comparativas da performance urbana ao longo do tempo, criando cenários para 5 e 10 anos. Avaliando as respectivas dinâmicas e riscos definindo tendências e gerando recomendações de urbanização, pois entendemos que este é o desafio do crescimento das cidades. Para tal, estamos criando indicadores de prosperidade urbano imobiliário.

Para vencer o desafio técnico, contamos com uma equipe de especialistas, no próximo capítulo detalhamos **“o como”** transformamos tecnicamente o **“sonho”** em **“realidade”**, ou seja, **“dados desestruturados e desconexos”** em **“insights relevantes para a melhor tomada de decisão”**.

Outro ponto de extrema importante no momento de estruturação da DataLand foi como nascer 100% conectado com todas as questões relacionadas à privacidade de dados, respeitando a individualidade e o consentimento de cada indivíduo, segundo os princípios da lei de proteção de dados.

No 3º capítulo tratamos da implantação e governança de dados da DataLand, pois para nós os dados devem ser tratados com transparência e responsabilidade.

Impactando no futuro das cidades

Coletamos dados nas mais diversas fontes, condicionadas ou relacionadas a Construção Civil ou aos resultados que desejamos atingir. Atualmente, nossos mecanismos de coleta atuam em fontes que disponibilizam dados de empresas, terrenos, regiões da cidade, dados socioeconômicos, populacionais, de segurança pública, clima e quaisquer outros que estejam de alguma forma relacionados aos desafios de negócio que estamos nos propondo resolver.

Por exemplo, podemos coletar e construir grandes bases de dados relativas a qualquer SQL (Setor Quadra e Lote) da cidade de São Paulo por meio de “processadores” específicos para aquele domínio (conjuntos de dados agregados e organizados que reproduzem uma necessidade de negócio). Nossas BUs (Business Units), trabalham com o conceito de

‘Data Capability’, que se refere à capacidade de ajustar um determinado conjunto de tecnologias para a criação de uma aplicação específica, resolvendo um problema de negócios. Assim, cada aplicação começa a ter a sua própria metodologia e tecnologia de manuseio e armazenamento de dados. Isso é o que se chama de ‘fit your data for your purpose’ (adequar o conjunto de dados ao seu propósito)”. Exemplificamos este conceito com a utilização da nossa aplicação de mapa, em que trabalhamos nosso streaming de dados com o Apache Beam e libs em Phyton para cálculos geoespaciais criando camadas proprietárias.

Os dados coletados são armazenados em nosso Data Lake, nos mais diversos formatos e estágios (stages). Efetuamos o armazenamento do dado no estágio raw, refletindo sua fonte, e a partir de então, nos vários



stages das suas transformações, até seu Domínio final.

A partir do Data Lake, um Cientista de Dados pode fazer análises exploratórias e gerar modelos de predição, clusterização como for necessário para atender as regras de negócio, gerando informações e insights de acordo com necessidades e domínios específicos.

O Data Lake é diferente do Data Warehouse, que se caracteriza por uma massa de dados já classificados e relacionados entre si. O conceito surgiu nos anos 80 para tentar organizar o Big Data interno das empresas.

Data Science é a disciplina responsável por classificar, categorizar e identificar os dados corretamente. Portanto, o profissional chamado Data Scientist (Cientista de Dados) é literalmente alguém que aplica conceitos e métodos científicos ao estudo de massas de dados, utilizando técnicas de Estatística e Matemática para explorar os dados e extrair informações e insights deles.

O acesso à informação e insights descobertos pelo Data Science, bem como sua visualização, pode ser feita das mais variadas maneiras. Podemos lançar mão de dashboards, gráficos, relatórios, renderização em sistemas de geoprocessamento, gráficos em grafos, entre outros, para uma tomada de decisão baseada em informações estruturadas.



Na figura acima, exemplificamos da coleta ao consumo do dado:



COMO TRATAMOS OS NOSSOS DADOS

Segundo os princípios da LGPD

A parte mais importante do trabalho da DataLand é a coleta de dados, pois é o nosso insumo e o que utilizamos para proporcionar soluções data driven para os nossos clientes. Como também tratamos e armazenamos uma quantidade imensa de informação, é nossa missão garantir que façamos isso em compliance com a LGPD, de forma transparente e respeitando a privacidade dos proprietário dos dados, e garantindo o cumprimento das obrigações legais.

As fontes de dados tratados pela DataLand são em grande parte fontes públicas, que dizem respeito à informações que são abertas em atendimento à LAI (Lei de Acesso à Informação), que determina que o poder público, por meio de suas diversas autarquias, órgãos, secretarias, dentre outros, devem prestar contas à população de certas informações para garantir o princípio da publicidade (tornar público) e de transparência para escrutínio do povo. Ou seja, a LAI incentiva que o poder público torne acessível a sua base de dados para a população, respeitando agora também a LGPD e também outros dados que são exceção à LAI.

Por exemplo, os municípios podem tornar públicos certas informações sobre sua base de cidadãos, mas devem garantir que, ao fazer isso, estão respeitando os princípios da LGPD, protegendo a privacidade dos indivíduos. Por premissa, a LAI já garante que dados sensíveis não podem ser divulgados. E nesse processo, deve também cumprir todas as funções de governança de dados junto aos proprietários dos dados conforme a LGPD.

O tratamento e uso desses dados também devem ser compatíveis com a finalidade e origem específicas para o qual o dado foi publicado em princípio. Na DataLand, olhamos para essas bases de dados abertos e escolhemos aqueles que fazem sentido serem tratados para gerar inteligência para os nossos produtos de Urban e AG Data Science.

LEGITIMIDADE DO TRATAMENTO



Portanto, aproveitamos a possibilidade da reutilização dos dados públicos para aquelas finalidades alinhadas ao nosso propósito, gerando conhecimento a partir deles.

E como trabalhamos o conceito de **“Privacy by Design”**? Nós olhamos durante a concepção dos nossos produtos/soluções para o propósito que queremos atingir. Alinhados com esse propósito, nós começamos então a procurar as fontes/origens de dados que nos permitirão atingir esse propósito, que no nosso caso é gerar inteligência para incorporação imobiliária e uso de terras no agronegócio. A partir da identificação das fontes de coleta, nós identificamos quais dados existem nessa base e como nossos algoritmos devem ser projetados para que as regras da LGPD

sejam cumpridas. Assim, nossa solução já nasce garantindo privacidade desde a concepção pela equipe de produto de software. Além de também garantir a segurança das informações sob a guarda da DataLand, impedindo vazamentos de qualquer natureza ou o uso indevido em desacordo com a finalidade dos dados.

Em termos técnicos, devem ser cumpridas as exigências da LGPD com relação à trilha de rastreabilidade dos dados - quando foi coletado, de onde, por quem, como - para que esse processo possa estar documentado e caso alguma autoridade solicite uma auditoria nesse processo, ele possa ser demonstrado, sem necessariamente divulgar o algoritmo, que é propriedade intelectual. E isso também permite provar que a análise tem legitimidade e cumpre os requisitos legais de tratamento.



E em se tratando de dados públicos, ainda existe o desafio de normalizar e compatibilizar fontes de dados a serem cruzados. Como cada órgão é responsável pelo seu PDA (Plano de Dados Abertos), ainda não existe uma padronização na disponibilização de dados. Cada cidade pode ter formatações ou maneiras diferentes de abrir seus dados, via APIs ou arquivos brutos. Além de poderem haver divergências entre dados similares, vindo de bases diferentes, visto que hoje as informações de bancos de dados públicos ainda não são totalmente conectadas. Por exemplo: se uma pessoa muda de nome, pode ser que encontremos nomes diferentes atrelados ao mesmo CPF em bases diferentes. Daí surge a necessidade de ter processos e algoritmos para cruzar essas informações. No entanto, mesmo identificando esses eventuais erros, é necessário garantir a integridade dos dados, ou seja, não é permitido alterar esses dados da fonte, mesmo que durante o tratamento seja inferida a informação correta ou retificada do mesmo.

Além das bases públicas, podemos produzir outras coletas em fontes particulares, mantendo os princípios da LGPD quanto ao consentimento na coleta, anonimização de dados e todas as restrições relacionadas aos dados sensíveis. Existem diversos mecanismos já praticados por empresa para essa coleta, e contanto que seja respeitada a privacidade dos usuários e a governança sobre os seus dados, é muito benéfica a reutilização dessas bases para o ambiente de negócios.



Conclusão

A DataLand, realmente tem como objetivo apoiar o desenvolvimento econômico e a melhoria da qualidade de vida das pessoas. Somos um exemplo de como aproveitar o acesso a dados abertos de forma legítima (em acordo com sua finalidade) e respeito privacidade do indivíduo para gerar valor para os nossos clientes.

Criamos uma relação de confiança legítima e geramos inteligência reutilizando os dados disponibilizados pela administração pública. Estamos trilhando o caminho que já vemos em cidades na Europa, que mantém essa relação com o uso de dados históricos em benefício da população e dos negócios.

E o que estamos desenvolvendo é um passo muito importante nesse sentido, para construirmos e cultivarmos de forma sustentável em todos

os aspectos, pois quanto menos risco de investimentos, melhor estamos gerindo os recursos naturais, desperdiçamos menos e possibilitamos soluções melhores de moradia, incentivamos cidades mais humanas e um abastecimento alimentar consciente e que respeite o meio ambiente.

Enxergamos as cidades e pessoas conectadas!! O ser humano tem uma necessidade inerente de ocupar o espaço construído e cultivar as terras. Se pudermos contribuir para fazê-lo de maneira a gerar menos impacto ambiental e mais prosperidade, não pouparemos esforços para isso.

Queremos que, com o nosso exemplo, outras empresas possam trilhar seu caminho para se adequarem à LGPD e percebam também a oportunidade de utilizar o poder dos dados a favor de soluções inovadoras para o Brasil e a nível global, respeitando o direito à privacidade de todos também.

Um ebook por



DATA LAND

LAND POWERED BY DATA SCIENCE

 @dataland-br

[ACESSE O SITE](#)