**1. Problem Framing & Stakeholder Analysis**

* **Define the "North Star Metric"**:  
  E.g., *“Reduce average commute time in urban areas by 20% using predictive data modeling.”*
* **Stakeholder Mapping**:  
  Identify key personas:
  + Commuters (end users)
  + City traffic control centers
  + Transit authorities
  + Logistics/fleet companies
  + Civic policymakers
* **Jobs-to-be-Done (JTBD)** analysis for each group.

**2. Opportunity Sizing & Impact Analysis**

* **Quantify the problem**:
  + Analyze open datasets (INRIX, TomTom, World Bank) to calculate cost of congestion in target cities.
  + Estimate potential emissions savings, time savings, productivity boosts.
* **TAM/SAM/SOM modeling**:
  + *TAM:* Global urban traffic market
  + *SAM:* Smart city analytics solutions
  + *SOM:* Pilot metro and urban partnerships

**3. Data Feasibility Study**

* **Evaluate Data Availability & Quality**:
  + Google Maps API (travel time, congestion status)
  + OpenWeatherMap API
  + Public transit and event feeds (GTFS, Ticketmaster, Eventbrite)
  + Historical traffic datasets (city-level or via Sidewalk Labs/open gov)
* **Schema and granularity check**:
  + Are travel times available at 5-min intervals?
  + Can routes be broken down into granular segments (polyline → road-level)?
  + What geo-spatial tags (lat/lng, geohash) are included?

**4. Technical Architecture Drafting**

* **First-pass system design (L0/L1 diagrams)**:
  + Create architecture diagrams like:
    - L0: User flow + data touchpoints
    - L1: High-level pipeline (Ingestion → ETL → ML → Dashboard)
* **Tech viability review**:
  + Does Kafka support target throughput for streaming travel updates?
  + Are Dataflow and BigQuery sufficient for latency-sensitive batch jobs?
  + How well can Vertex AI handle time-series congestion prediction?

**5. Legal, Ethical & Privacy Review**

* **Compliance Check**:
  + Is location data anonymized?
  + Do third-party APIs (e.g., Google Maps) allow downstream ML training?
  + Assess GDPR/CCPA compliance if storing trip data.
* **Bias/Risk Evaluation**:
  + Could the model unintentionally favor wealthier areas with cleaner data?
  + Are there blind spots in event or weather coverage?

**6. Pilot City Scoping**

* Shortlist cities based on:
  + Data availability
  + Willingness of civic authorities to collaborate
  + Volume of commuters + congestion hotspots
  + Google’s existing infrastructure (Fiber, GCP zones, Maps ops)

**7. Early Prototyping + Stakeholder Demos**

* **Design a prototype with real data**:
  + Simulate streaming ingestion + travel time prediction
  + Build a basic Streamlit dashboard + heatmap
  + Show “before vs after” simulated congestion
* **Run internal UX sessions** with Google UX Research team
* Prepare demo for civic partners or “Moonshot Review Board”

**8. Resource Planning & Budget Forecasting**

* **Estimate cloud compute/storage/APIs per month**
* **Build a cross-functional team**:
  + Data Engineers
  + ML Engineers
  + Geo specialists
  + Cloud Architects
  + Policy analysts
  + Product leads
* Align with GCP, Geo (Maps), and Sidewalk Labs teams

**9. Define Go/No-Go Criteria**

* Data quality >= 85% across major corridors
* Predictive accuracy (MAE < 5 mins on congestion model)
* Pilot city agreement in place
* MVP latency targets (ETL < 15 min, prediction API < 500ms)

**10. Executive Review / Greenlight**

* Present findings to SVP-level stakeholders
* Create a one-pager, visual walkthrough, and Gantt-based execution roadmap
* Secure seed budget or formal transfer to Moonshot (X) or Sidewalk Labs