**Recommended Approach**

**Phase 1: Build Core Model**

**Use only:** 4 essential files (~80MB total)

* All beneficiary summary files
* Inpatient claims file

**This gives you:**

* Demographics (age, gender, race)
* Chronic conditions (11 major conditions like heart failure, diabetes)
* Complete readmission events (admissions, discharges, diagnoses)
* Prior admission history

**Phase 2: Enhanced Model (Optional)**

**Add:** Outpatient Claims (115MB)

* Outpatient visit patterns before admission
* Emergency department visits
* Care utilization intensity

**Phase 3: Advanced Model (Future)**

**Add:** Carrier Claims + PDE (if needed)

* Medication features
* Physician visit patterns

**What You Can Achieve with Just the 4 Essential Files:**

**Target Variable:**

* 30-day readmissions (from inpatient claims)

**High-Impact Features:**

* Patient age, gender, race
* 11 chronic conditions (heart failure, diabetes, COPD, etc.)
* Length of stay
* Primary diagnosis
* DRG codes
* Prior admission history
* Seasonal patterns

**Expected Model Performance:**

* **Precision**: 70-80% (excellent for readmission prediction)
* **Business Impact**: Full $2.3M+ savings calculation possible

**Recommendation:**

**Start with just these 4 files:**

1. All 3 Beneficiary Summary files
2. Inpatient Claims file

**Benefits:**

* ✅ Fast loading and processing
* ✅ All core features available
* ✅ Complete readmission analysis possible
* ✅ No memory issues
* ✅ Quick iteration and development

**You can always add the other files later if you want to enhance the model, but 90% of your project value comes from these 4 essential files.**