=== Метрика: Euclidean, K=2 ===

Центроиды:

Центроид 1: x=29.9873, y=32.2729

Центроид 2: x=63.1095, y=68.5856

Кластер 1: Дисперсия=152.0868, Радиус=30.3119

Кластер 2: Дисперсия=650.7025, Радиус=55.9603

Суммарное квадратичное отклонение=144351.9552

Силуэтный коэффициент=0.5146

=== Метрика: Euclidean, K=3 ===

Центроиды:

Центроид 1: x=29.3866, y=32.1968

Центроид 2: x=78.9887, y=52.2033

Центроид 3: x=47.7925, y=83.6880

Кластер 1: Дисперсия=136.3936, Радиус=22.9474

Кластер 2: Дисперсия=200.2339, Радиус=37.3771

Кластер 3: Дисперсия=128.8760, Радиус=25.0732

Суммарное квадратичное отклонение=46407.6331

Силуэтный коэффициент=0.6411

=== Метрика: Euclidean, K=4 ===

Центроиды:

Центроид 1: x=69.5673, y=58.4838

Центроид 2: x=86.1146, y=48.2593

Центроид 3: x=46.8494, y=84.6630

Центроид 4: x=29.3866, y=32.1968

Кластер 1: Дисперсия=154.3560, Радиус=25.4814

Кластер 2: Дисперсия=125.1572, Радиус=30.4016

Кластер 3: Дисперсия=105.1550, Радиус=20.8190

Кластер 4: Дисперсия=136.3936, Радиус=22.9474

Суммарное квадратичное отклонение=38318.1232

Силуэтный коэффициент=0.3940

=== Метрика: Manhattan, K=2 ===

Центроиды:

Центроид 1: x=38.4694, y=58.0006

Центроид 2: x=79.0104, y=52.6632

Кластер 1: Дисперсия=1395.4867, Радиус=59.2068

Кластер 2: Дисперсия=332.6197, Радиус=38.6374

Суммарное квадратичное отклонение=313422.1742

Силуэтный коэффициент=0.3925

=== Метрика: Manhattan, K=3 ===

Центроиды:

Центроид 1: x=29.6916, y=32.2224

Центроид 2: x=79.1824, y=52.3829

Центроид 3: x=47.7925, y=83.6880

Кластер 1: Дисперсия=214.4727, Радиус=33.0694

Кластер 2: Дисперсия=308.5134, Радиус=38.1851

Кластер 3: Дисперсия=199.6333, Радиус=29.6678

Суммарное квадратичное отклонение=71950.1376

Силуэтный коэффициент=0.6352

=== Метрика: Manhattan, K=4 ===

Центроиды:

Центроид 1: x=70.5800, y=54.7557

Центроид 2: x=87.9559, y=50.8763

Центроид 3: x=47.1150, y=84.1376

Центроид 4: x=29.3866, y=32.1968

Кластер 1: Дисперсия=234.9168, Радиус=31.2270

Кластер 2: Дисперсия=207.6504, Радиус=28.8961

Кластер 3: Дисперсия=177.5062, Радиус=27.0527

Кластер 4: Дисперсия=205.6289, Радиус=29.5087

Суммарное квадратичное отклонение=60665.7556

Силуэтный коэффициент=0.4069

=== Метрика: Chebyshev, K=2 ===

Центроиды:

Центроид 1: x=29.9873, y=32.2729

Центроид 2: x=63.1095, y=68.5856

Кластер 1: Дисперсия=134.3803, Радиус=30.2074

Кластер 2: Дисперсия=484.1028, Радиус=53.2470

Суммарное квадратичное отклонение=109559.1488

Силуэтный коэффициент=0.5146

=== Метрика: Chebyshev, K=3 ===

Центроиды:

Центроид 1: x=29.3866, y=32.1968

Центроид 2: x=78.9887, y=52.2033

Центроид 3: x=47.7925, y=83.6880

Кластер 1: Дисперсия=118.6653, Радиус=22.3872

Кластер 2: Дисперсия=165.3387, Радиус=37.3678

Кластер 3: Дисперсия=109.5935, Радиус=24.5440

Суммарное квадратичное отклонение=39248.2613

Силуэтный коэффициент=0.6411

=== Метрика: Chebyshev, K=4 ===

Центроиды:

Центроид 1: x=73.7686, y=47.2068

Центроид 2: x=86.3556, y=57.5286

Центроид 3: x=48.2520, y=83.0821

Центроид 4: x=29.3866, y=32.1968

Кластер 1: Дисперсия=85.2402, Радиус=17.8219

Кластер 2: Дисперсия=131.3056, Радиус=30.0009

Кластер 3: Дисперсия=114.7125, Радиус=22.2577

Кластер 4: Дисперсия=118.6653, Радиус=22.3872

Суммарное квадратичное отклонение=33989.9752

Силуэтный коэффициент=0.3878

**Анализ и выводы по кластеризации**

Данные содержат результаты кластеризации с использованием трех метрик расстояния: **Euclidean**, **Manhattan**, **Chebyshev**, при разном числе кластеров KKK. Оценены следующие показатели: центроиды кластеров, дисперсия, радиус, суммарное квадратичное отклонение, силуэтный коэффициент.

**1. Выбор оптимального числа кластеров KKK:**

1. **Силуэтный коэффициент (Silhouette Coefficient):**
   * Максимальное значение этого показателя указывает на наиболее качественную кластеризацию.
   * При всех трех метриках для K=3K=3K=3 достигается наивысший силуэтный коэффициент (≈ 0.64). Это говорит о том, что данные хорошо разделяются на 3 группы.
2. **Суммарное квадратичное отклонение (Sum of Squared Deviation):**
   * С ростом KKK суммарное квадратичное отклонение уменьшается, что указывает на снижение внутрикластерной вариации.
   * Однако K=4K=4K=4 показывает значительно меньшее улучшение в сравнении с K=3K=3K=3, что соответствует правилу "локтя" (дальнейшее увеличение KKK не оправдано).

Вывод: K=3K=3K=3 является оптимальным выбором.

**2. Сравнение метрик расстояния:**

1. **Euclidean:**
   * Показывает наилучшие результаты по качеству кластеризации (суммарное квадратичное отклонение = 46,407.6331, силуэтный коэффициент = 0.6411 при K=3K=3K=3).
   * Дисперсии и радиусы кластеров сбалансированы, что указывает на стабильное распределение точек вокруг центроидов.
2. **Manhattan:**
   * Второй по качеству метрики, при K=3K=3K=3 силуэтный коэффициент составляет 0.6352, а суммарное квадратичное отклонение = 71,950.1376.
   * Радиусы кластеров выше, чем у Euclidean, что может указывать на более вытянутую форму кластеров.
3. **Chebyshev:**
   * Наименьшее суммарное квадратичное отклонение (39,248.2613 при K=3K=3K=3), но более низкий силуэтный коэффициент (0.6411).
   * Кластеры имеют меньшие радиусы и дисперсии, но могут быть менее четко разделены, чем при Euclidean.

Вывод: **Euclidean** метрика является предпочтительной из-за наиболее высокого качества кластеризации.

**3. Интерпретация кластеров для K=3K=3K=3:**

1. **Кластер 1 (на примере Euclidean):**
   * Центроид: x=29.39,y=32.20x=29.39, y=32.20x=29.39,y=32.20.
   * Дисперсия: 136.39, Радиус: 22.95.
   * Характеризует компактную группу точек, наиболее удаленных от других кластеров.
2. **Кластер 2:**
   * Центроид: x=78.99,y=52.20x=78.99, y=52.20x=78.99,y=52.20.
   * Дисперсия: 200.23, Радиус: 37.38.
   * Наиболее вытянутый кластер, указывающий на группу точек с большим разбросом.
3. **Кластер 3:**
   * Центроид: x=47.79,y=83.69x=47.79, y=83.69x=47.79,y=83.69.
   * Дисперсия: 128.88, Радиус: 25.07.
   * Кластер с меньшим радиусом, точки более равномерно распределены вокруг центроида.

**4. Итоговые рекомендации:**

1. **Оптимальные параметры:**
   * Использовать K=3K=3K=3 и метрику Euclidean для кластеризации данных.
   * Это обеспечивает хорошее качество кластеризации, сбалансированные радиусы и минимальные отклонения.
2. **Дополнительные действия:**
   * Для проверки устойчивости модели провести анализ с другими методами оценки качества кластеризации (например, индекс Дэвиса-Болдина).
   * Рассмотреть использование визуализации кластеров (например, PCA или t-SNE) для лучшего понимания структуры данных.
3. **Особенности интерпретации:**
   * Учесть, что радиусы кластеров зависят от используемой метрики. Для более вытянутых или несимметричных данных Manhattan или Chebyshev могут быть полезны.

Таким образом, выбранный метод и параметры (Euclidean, K=3K=3K=3) обеспечивают наилучшее разделение данных на группы.